Journal of
Experimental Psychology

VOL. 30, NO. 3 MARCH, 1942

INTRA-LIST GENERALIZATION AS A FACTOR IN VERBAL LEARNING [1]

BY ELEANOR J. GIBSON

Smith College

INTRODUCTION

The present experiment is concerned with the role played by similarity or homogeneity of items in verbal learning. It is related, therefore, to the problem of transfer in general. But the specific problem to be dealt with here is the effect of mutual similarity of items within a *single* list or task. What is the effect upon learning of increasing the similarity of such units? Previous studies of the interrelation of items within a learning task have almost all been concerned with serial position effects. So far as the writer is aware, a study by von Restorff (**13**) is the only previous one concerned with the effect of intraserial similarities on learning efficiency. Von Restorff showed that isolated items within an otherwise homogeneous list were recalled better than the same items in surroundings homogeneous with themselves. She interpreted this phenomenon in terms of the Gestalt laws of perception applied to 'trace fields.' The present approach differs from hers in systematic background and in its emphasis on *discriminability*, rather than similarity considered as formal likeness. The establishment of differentiation between items or units of a task is emphasized as an essential factor in verbal learning. Relative indistinguishability of such items at the beginning of a learning task necessitates a process of differentiation over and above the forming of connections which the instructions call for. Such a lack of identifiability of items is frequent in the usual verbal learning situation, especially since most experimental conditions exclude any

[1] This experiment is part of a dissertation presented to the faculty of Yale University in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Professor Clark Hull for many valuable criticisms.

direct perceptual comparisons. The resulting functional equivalence of stimuli appears to be analogous to Pavlovian stimulus-generalization,[2] and this analogy forms the basis of a theoretical structure relating verbal learning to the conditioned response concepts of generalization and differentiation (3).[3]

The hypothesis assumes that generalization may occur in different degrees among the stimulus items of verbal lists, so that a response learned to a given stimulus item may tend to occur to others in the list as well. For convenience of analysis, the paired associates method will be referred to throughout the description. Stimulus-generalization may occur equally well with other methods, but isolation of its effects is more difficult in a list where an item may function as both a stimulus and a response. Figure 1 represents list of paired

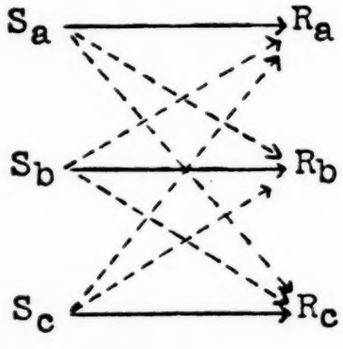

FIG. 1

associates in which generalization is assumed to be taking place during learning. The first members of the pairs are assumed to function as stimuli. The stimulus items S_a, S_b, and S_c tend to call up one anothers' responses as well as their own. Generalization tendencies (represented by broken arrows) will tend to block the right responses (solid arrows) in proportion to the strength of the tendencies. In order to reduce the strength of the generalizing tendencies, differential reinforcement must be applied—in other words, differentiation between the stimulus items must be set up through practice. It is assumed that varying degrees of generalization may occur between one item and others in a list, and consequently that lists may vary as to the average strength of generalization occurring in a list. The necessity for greater differentiation will then serve to make the learning more difficult in lists where a high degree of generalization

[2] Hull (8) has made a detailed analysis of the relationship between stimulus equivalence and generalization.

[3] Generalization and differentiation are both defined behaviorally: generalization as the tendency for a response learned to one stimulus to occur when another stimulus is presented with which it has not previously been associated; differentiation as a progressive decrease in generalization resulting from reinforced practice with one S-R combination and unreinforced presentation of the other stimulus.

occurs than in ones where less generalization occurs. The analogy with conditioning situations suggests also that differentiation may lapse over a period of time, leading to a 'spontaneous recovery' of generalization (**6**); retention of a list would thus be poorer, the greater the degree of intra-list generalization, when a sufficient interval had elapsed since learning. In the light of the hypothesis, then, it would be expected that high-generalization lists should be harder to learn than low-generalization lists; and that the high-generalization lists should be less well retained after a period of time. It is these expectations which the present experiment is designed to test.

The hypothesis also makes an assumption regarding the course of generalization during a learning task: that generalization *increases* during the early stages of practice, and then *declines* as differential reinforcement of right and wrong responses has an opportunity to occur. It assumes, furthermore, that more differential reinforcement is required for any particular association, the greater the strength of the competing generalizing tendencies. The experiment provides an opportunity for checking these assumptions.

Method

Standardization of Degree of Generalization.—The basic plan of the experiment was a comparison of the ease of learning different lists of paired associates which varied in the degree to which their stimulus members generalized with one another. Relative retention of the lists was compared by securing recall and relearning scores one day after learning. Since the independent variable was to be degree of generalization of stimulus members within the list, it was essential that an objective measure of degree of generalization be secured. A preliminary experiment designed to meet this requirement has been described by the author (**4**). The general technique was as follows. A series of *standard* nonsense forms was selected, to be used as first members or cues in paired associates learning. Two *variations* from each of these forms were also employed, constituting two stages of increasing dissimilarity. Twenty-four hours after subjects had learned the list of standard forms paired with words, a recall series was given them, in which each cue form was either the original standard, or a variation substituted for the standard. The measure of generalization was taken as the extent to which the subjects responded to the variation *as if it were the standard.* After rotating the forms in recall lists with various groups of subjects, the experiment yielded a measure of frequency of recall of the appropriate word for each standard form and for each variation on the standard. Hence there were established three classes of forms, corresponding to three degrees of stimulus-equivalence. Class I consisted of the identical standard forms, which were most frequently responded to (84.5 percent); Class II contained those variations which produced the highest frequency of response-as-if-to-the-standard (41.1 percent); and Class III contained those variations less frequently responded to (9.7 percent).[4] Individual percents are presented with the forms in (**4**). There was a negligible tendency for one standard form to call forth the response appropriate to any other standard. The subjects realized that the recall list was not the intact original list, so that when a variation recalled a standard response, it was functioning as an *equivalent* stimulus, not merely as a recognizably similar one.

Construction of Lists.—Lists of these forms paired with nonsense syllables were then constructed for the present experiment, so as to yield one list having a low degree of generalization among its stimulus items, and three lists having a high degree of internal stimulus-generalization. Figure 2 is a schema of the stimulus members of the four lists. The low-generalization list (Series I) contained 12 standard forms. Each high-generalization list contained four groups of gen-

[4] Actually, a fourth class was also yielded, which gave an almost negligible frequency of response, but it was not used in the present experiment.

eralizing forms, a standard and its two variations. In Series II *A*, for instance, the standards *A*, *B*, *C*, and *D* are present, and the other eight forms consist of their several variations. Since the variations have been shown to have a significant tendency to call forth the responses of their standards, these lists may be characterized as having high internal generalization as compared with Series I. It was essential to have three lists in Series II, so that each standard might appear in a high-generalization list as well as a low one, thereby making comparison of individual forms possible. The same 12 response-syllables were used in all four lists, and a standard form was paired with the same syllable in both lists in which it appeared. Position in the list order was equalized as well. Since it was desired to vary the order of presentation in successive learning

Low-Generalization List	High-Generalization Lists		
Series I	Series II *A*	Series II *B*	Series II *C*
A	*A*	*E*	*I*
B	A_1	E_1	I_1
C	A_2	E_2	I_2
D	*B*	*F*	*J*
E	B_1	F_1	J_1
F	B_2	F_2	J_2
G	*C*	*G*	*K*
H	C_1	G_1	K_1
I	C_2	G_2	K_2
J	*D*	*H*	*L*
K	D_1	H_1	L_1
L	D_2	H_2	L_2

FIG. 2. Diagram of stimulus members of the four lists. Letters represent forms arranged so as to yield one low-generalization list, and three high-generalization lists. Letters with subscripts represent variations; others represent standard forms.

trials, six random orders for the paired items were arranged, and a given standard which occupied position 1 in random arrangement *a* of the low-generalization list also occupied position 1 in random arrangement *a* of the high-generalization list. The standard items could thus be compared independently, the only variable being the similarity or difference of the stimulus members which surrounded them during learning. The lists were constructed by pasting mimeographed copies of the forms on heavy drawing paper and typing the syllables beside them in capital letters.

Procedure.—The material was presented to the subjects on an electrically driven Chicago-type memory drum. They learned individually by the method of right associates. That is, each pair was exposed for 2 seconds in a learning trial (one complete presentation of the list); a test trial followed in which the forms alone appeared for 2 seconds, and the subject attempted to recall the correct associate. The order was changed between learning trials as well as before each test trial. Learning and test trials alternated until the subject could give all 12 correct associates. All the subjects' responses were recorded by the experimenter.

The following instructions were given the subject at the beginning of the experiment:

> You will be shown a group of forms, each one paired with a nonsense syllable. Study these pairs so that when a form is shown alone, you can recall the appropriate syllable. You will be shown only one pair at a time. Do not try to learn these pairs in any particular order, because the order will be changed every time. The point is to associate a particular syllable with the form with which it always appears.
>
> After every presentation of the forms paired with syllables, you will be shown the forms by themselves in order to see whether you remember the appropriate syllable or not. If you do remember it, speak it aloud. We shall continue until you recall all the syllables correctly in one trial.

The subject returned at the same hour the following day, at which time he was given a test trial, and then relearned the list. No subject was used more than once, since some of the forms

appeared in both series. There were 14 subjects in each of the four groups. They were Smith college students majoring in psychology and students from the Smith College School for Social Work.

Results

Learning and Retention of Low- and High-Generalization Lists.—The comparative data for learning, recall, and relearning of the low- and high-generalization lists are presented in Table I.

TABLE I

Comparison of Ease of Learning and Degree of Retention of High- and Low-Generalization Lists

	Low-Generalization List	High-Generalization Lists		
	Series I	Series II *A*	Series II *B*	Series II *C*
Mean Trials to Learn.....	8.86	22.29	14.36	22.79
Mean Trials to Relearn...	2.00	2.64	2.79	2.79
Mean Number Recalled...	9.31	8.86	8.50	8.71

The low-generalization list (Series I) required only 8.86 trials to be learned to a criterion of one perfect trial, on the average, while the three high-generalization lists (Series II *A*, II *B*, and II *C*) required respectively 22.29, 14.36, and 22.79 trials.[5] The data thus show that lists having a relatively high degree of internal stimulus-generalization are much harder to learn than those having a relatively low degree, as the hypothesis would predict. Furthermore, in spite of the much greater number of learning repetitions given the high-generalization lists, they are not so well retained as the low-generalization list. All three of the Series II lists required a slightly greater number of relearning trials than the Series I list; and in all three, fewer responses were correctly recalled after 24 hours. These differences in retention are not statistically reliable, but they occur consistently in the three lists, by both measures. Even this small difference is rather striking, considering the great opportunities for overlearning in Series II. The greater the overlearning, the longer spontaneous recovery of generalization should be postponed, according to the hypothesis, so that a recall test later than 24 hours might have shown even a greater difference.

Individual Comparison of Pairs of Items in Low- and High-Generalization Lists.—The comparison of the two types of list which has just been made might conceivably be objected to on the ground that certain stimulus items occur in the Series II lists which do not occur in the Series I list, and that some of these stimuli might be intrinsi-

[5] The critical ratios (D/σ diff.) of the differences between Series I and Series II *A*, II *B*, and II *C* are respectively 4.53, 3.43, and 3.93.

cally harder to associate with the responses paired with them than the forms of Series I. It is most unlikely that such a factor could be responsible for the large differences between the series. But a comparison can be made which rules it out beyond a doubt. The standard forms and their responses of Series I can be compared individually with the same forms and responses as members of Series II, the sole variable being the strength of the tendency to generalize with other members of the list in the two cases. Table II shows the number of

TABLE II

COMPARISON OF INDIVIDUAL PAIRS IN HIGH- AND LOW-GENERALIZATION LISTS

No. of Item	Mean No. of Repetitions to First Correct Response		Mean No. of Repetitions to Second Correct Response		Mean No. of Repetitions to Relearn	
	I In Low-Generalization List	II In High-Generalization List	I In Low-Generalization List	II In High-Generalization List	I In Low-Generalization List	II In High-Generalization List
1	5.14	10.86	6.15	13.07	.50	.21
2	3.00	9.00	4.79	11.17	.50	1.21
3	4.86	7.57	5.62	8.84	.57	.64
4	2.79	6.93	3.79	8.92	.21	.36
5	2.64	3.64	3.46	5.23	.21	.14
6	2.93	6.57	4.43	9.57	.07	.14
7	2.50	10.93	3.86	14.36	.07	.29
8	3.79	9.14	4.62	12.23	.64	1.21
9	4.50	11.21	5.31	12.69	.71	.57
10	2.07	3.79	3.07	5.86	.14	.50
11	4.00	8.14	4.62	10.14	.43	.36
12	3.07	5.50	3.92	9.79	.36	.50

repetitions required before the correct response was *first made* to a particular stimulus item when it was a member of Series I and when it was a member of Series II. For every pair, a greater number of presentations was required to reach the first correct response in Series II. Some pairs required three times as long to be learned.

It is conceivable that the first correct response might be in the nature of an accident—that is, no more a real criterion of learning than a generalized response would be. Therefore, the number of trials required before reaching the *second* correct response was also calculated for each item. Again, every item required many more trials in Series II, the high-generalization series, the differences being even greater by this measure. The consistency of the data supports beyond a doubt the conclusion that it is harder to learn a particular form-syllable pair when it is surrounded by other pairs having generalizing stimulus items.

In Table II are also included the mean number of relearning presentations required to recall each response correctly under the two conditions. These means are mostly under 1, because an item was

recalled by many subjects before any relearning exposures were given—that is, in the first recall test. In 8 out of the 12 cases, the mean is higher in the high-generalization series. That the difference between the two series is not greater when measured by a relearning score for each item is undoubtedly due to great overlearning in Series II.

Errors of Overt Generalization.—The hypothesis which has been described, as well as the logic of the present experiment, demands that a greater number of errors of overt generalization (confusion errors) occur in Series II than in Series I. By a generalization error is meant a wrong response which is actually the appropriate response to another stimulus item in the list. The hypothesis assumes that such wrong responses occur by virtue of generalization between the stimulus members. The total number of such errors during learning, and during recall and relearning for each of the series is presented in Table III. Many more such errors occurred in each of the Series

TABLE III

TOTAL NUMBER OF ERRORS OF OVERT GENERALIZATION

	Series I	Series II *A*	Series II *B*	Series II *C*
In Learning	43	293	139	323
In Recall and Relearning	8	50	29	27

II lists for both learning and retention than in Series I. Series II *B* had fewer generalization errors than the other two high-generalization lists. It may be noted, and it tends to verify our hypothesis, that this list also required fewer learning repetitions than the other two (14.36 trials as compared with 22.29 and 22.79).

It will be remembered that in each of the high-generalization lists, four groups of forms were included. Each group was made up of a standard and two variations from the standard which had been shown to generalize with it. But between the groups themselves there was no relationship—no more, at least, than between the stimulus members of Series I. It is thus crucial to see whether most of the errors of overt generalization were responses paired with stimulus items belonging to the same group as the item which elicited them. When all such generalized responses to similar items are taken together, they form the bulk of the confusion errors in each of the three lists—in Series II *A*, 88 percent, in Series II *B*, 80 percent and in Series II *C*, 80 percent. Thus, there is no doubt that intra-list generalization occurs most frequently between 'similar' stimulus items.

It is possible to analyze these errors still further. Within each group of three similar items in a high-generalization list, two degrees

of generalization may be distinguished, since in the standardization experiment one of the variations elicited a relatively high frequency of response-as-if-to-the-standard, and the other a relatively low one. Cases of generalization between the standard and the first variation will be called Type 1 errors; generalization between a standard and the second variation will be called Type 2 errors. Cases of generalization between the two variations themselves cannot be classified as either Type 1 or Type 2 errors, since neither has been a standard for the other. Therefore, they will be designated Type X errors. Cases of generalization between items belonging to different groups will be called Type 3 errors. It should be noted that these four

TABLE IV

CLASSIFICATION OF ERRORS OF OVERT GENERALIZATION

	Series II *A*	Series II *B*	Series II *C*	Percent of Total
During Learning				
Type 1	169	75	70	42%
Type 2	42	25	95	21%
Type X	46	11	89	19%
Type 3	36	28	69	18%
				100%
During Recall and Relearning				
Type 1	30	16	7	50%
Type 2	8	6	12	25%
Type X	7	2	6	14%
Type 3	5	5	2	11%
				100%

types exhaust all the possibilities of errors except for invented syllables. Table IV shows the relative frequency of the four types of error. As the hypothesis demands, more Type 1 errors occur than any other kind. Twice as many Type 1 errors occur as Type 2 errors, during both learning and relearning of the lists. We can have no particular expectation as regards Type X errors, except that they should be relatively more frequent than the Type 3 errors, as they are. It is really a remarkable verification of the theoretical prediction that Type 3 errors should be as infrequent as they are, since there are 9 times as many opportunities for these errors to occur as for the other types, and the frequencies given in the table might be divided by 9 with justification. It may be concluded that the frequency of errors of Types 1, 2 and 3 constitutes a gradient which fulfills completely the theoretical expectation.

When the three sub-series are examined separately, it will be observed that the consistency of the results is slightly impaired by Series II *C*, where Type 2 errors were more frequent than Type 1 errors. This is probably related to the fact that the two standard forms whose variations yielded the least difference in generalizing tendency happened to be included in this list. The other two sub-series bear out the trends shown in the totals.

Rate of Differentiation for Different Degrees of Generalization.—It is assumed by the hypothesis which is under scrutiny that the amount of reinforcement required to reduce a generalizing tendency to a given strength will increase with increasing strength of the generalizing tendency. It is possible under the conditions of this experiment to put to test this assumption as regards intra-list generalization (the assumption has been verified in Pavlov's laboratory for generalization of conditioned reflexes). Four types of overt generalization, described above, occurred in this experiment. If the assumption is correct, the type which occurred least frequently should have been the first to disappear (to be reduced below threshold

TABLE V

DISTRIBUTION OF THREE TYPES OF GENERALIZATION IN HALVES OF THE LEARNING PROCESS

	Type 1		Type 2		Type 3	
	First Half	Second Half	First Half	Second Half	First Half	Second Half
Series II *A*	79	89	26	16	28	8
Series II *B*	40	35	15	10	24	4
Series II *C*	42	28	51	44	51	18
Total	161	152	92	70	103	30
Percent of Total	51%	49%	57%	43%	77%	23%

strength), and so on. In other words, stimuli of low generalizing tendency should be effective in producing generalization with one another in the early stages of learning, but not in the later stages. Table V shows the distribution of three of the types of generalization in halves of the learning process. Type X is not included, since the degree of generalization was not known accurately in this case. Generalization of Type 1 (highest degree of generalization) occurred almost equally in both halves of the learning process—51 percent of the errors occurred in the first half, and 49 percent occurred in the second half. Generalization of Type 2 (second highest degree of generalization) occurred more frequently in the first half—57 percent to 43 percent. Generalization of Type 3 (lowest degree of generalization) occurred almost entirely during the first half of the learning process—77 percent as compared with 23 percent. The results for

the three series taken separately are in each case quite consistent with the totals, the difference in frequency between first half and second half increasing as degree of generalization decreases. It can therefore be concluded that differentiation is reached sooner for stimulus items which have a low tendency to generalize, while more differential reinforcement is required for items with stronger tendencies to generalize. This fact is important also because it means that the absolute degree of generalization which is revealed in a test will vary with the degree of learning previously achieved. This tendency is demonstrated with greater precision in the Vincent curves below.

Vincent Curves of Generalization.—It was necessary, in the development of the hypothesis presented, to make some assumption with respect to the nature of the curve of generalization during verbal learning. The degree of generalization of conditioned reflexes as dependent upon different amounts of reinforcement has been studied by Pavlov (**12**), by Beritoff (**1**), and particularly by Hovland (**7**). Although their experiments were conducted under rather different conditions, all three investigators reported that generalization at first increased and subsequently decreased as more reinforcements were given the conditioned stimulus. By analogy with this fact, it was tentatively assumed that generalization at first increased to a peak and subsequently diminished as progressive learning trials were administered during learning of verbal lists. In the present experiment, an opportunity arose to check this assumption, since it was possible to plot the degree of overt generalization which occurred at various stages of the learning process. Although there undoubtedly were frequent tendencies to generalize which did not materialize in overt errors,[6] the number of cases of overt generalization may be taken as a good indicator of the strength of the generalizing tendencies. Since a significant curve could only be obtained by pooling the results of the subjects, and since the number of learning repetitions differed widely from subject to subject, some method of securing equal stages of learning with variable numbers of trials was required. Vincent's method as modified by Hunter (**9**) was adopted.

In Fig. 3 is plotted the curve of generalization during successive tenths of the learning process, in terms of mean number of generalized responses. In the curve are pooled the data for Series II *A*, II *B*, and II *C*. The curve of learning (right responses) is plotted on the same axes for purposes of comparison, with the units on the ordinate multiplied by 10. It may be seen that the assumption regarding the course of generalization is borne out, since the curve rises rapidly

[6] Subjects frequently reported a conflict between a generalized response and the right response after an omission had occurred.

during the first three tenths of the learning, and then falls slowly during the last seven tenths. There is a very slow fall from the third to the eighth tenth, which should be significant in making predictions

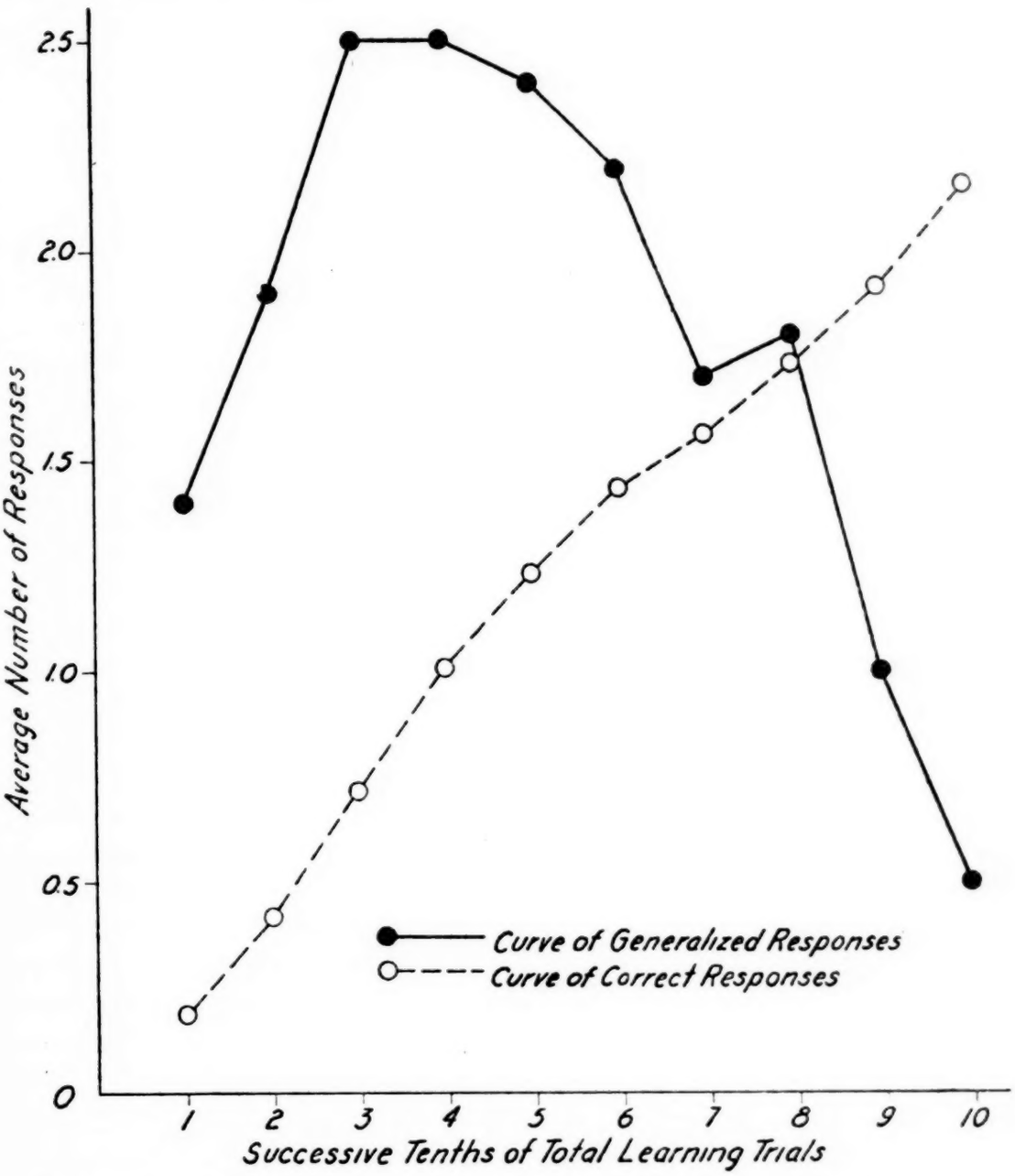

FIG. 3. Curve of generalization during successive tenths of learning for Series II *A*, *B*, and *C*. The learning curve (curve of right responses) has been plotted on the same axes for purposes of comparison. The units on the ordinate must be multiplied by 10 in order to read the learning curve.

having reference to degree of learning. The curves of generalization for Series II *A*, II *B*, and II *C*, when drawn separately, resemble the composite curve in every important respect.

The learning curve in Fig. 3 is interesting when studied in connection with the curve of generalization. The two curves rise to-

gether through the third tenth of the learning process, but from about the fifth tenth on, they are reciprocals of one another. A rise in generalization is expected during the first three tenths of the process, since there can obviously be no overt generalization when there are no excitatory tendencies whatsoever; and in the last third, as differentiation increases rapidly, so also should the number of right responses. One might be tempted to ascribe the shape of the curve of generalization to the same causes to which Pavlov imputed the initial increase and eventual decrease of generalization of conditioned reflexes. But it is necessary to consider first the experimental conditions in the two cases. Pavlov reports (**12**, p. 118) that an initial increase and eventually a decrease in generalization occurred when the effect of a closely allied stimulus was tried "after a conditioned reflex to a definite stimulus had been firmly established." In the present experiment, on the other hand, no learning whatsoever had taken place before the opportunity for generalization was offered. The conditions of the two experiments, then, are by no means parallel. Pavlov explained the initial rise of generalization as due to the falling off of external inhibition produced by a 'temporary investigatory reflex,' and the final decrease as due to the 'development of internal inhibition.' It is obvious that the rise in the curve obtained in the present experiment cannot be attributed to falling off of external inhibition of otherwise available excitatory tendencies, for no excitatory tendencies would be originally available. It seems more likely that a simpler explanation is adequate, namely that a minimum number of repetitions of each pair is necessary to produce excitatory tendencies of threshold strength—generalized ones as well as correct ones—and that while the minimum number for the list as a whole is being reached, generalized responses as well as right ones will increase. But when this point is reached, differential reinforcement can occur effectively, so that generalized responses will be on the wane, while right responses continue to increase. This is in keeping with the early parallel rise in the two curves and with their final reciprocal relationship.

The shape of the learning curve is interesting. Although the curve shows a tendency toward negative acceleration between the fourth and the seventh tenth, the rise from the seventh tenth on is almost linear. Probably this final more rapid rise may be identified with the so-called end-spurt obtained in curves of memorization when they are averaged by the Vincent-Kjerstad method (**10**).

Discussion

The results of the present experiment are in perfect accord with the hypothesis which was briefly outlined in the introduction. With

the results of a previous experiment (**4**), they confirm the suggestion that generalization frequently occurs between items of a list, and that as a consequence differentiation of the items is an important feature of verbal learning. In other words, verbal learning is not merely a matter of establishing bonds or connections between one item and the next, or one member and its associate. The *alikeness* of the items, their functional equivalence at the start, is responsible for tendencies toward generalization which must be overcome. Each item must become unique, discriminable from the others, in order that correct linkages can be formed. In the early stages of learning, we have shown, generalization increases. This is undoubtedly because linkages begin to be formed and strengthened as soon as practice begins, whereas the real process of differential reinforcement does not start until right and wrong responses are actually made by the subject. The fact that the curve of generalization *rises* for the first three-tenths of the learning process is itself an indication that differential reinforcement is the process whereby uniqueness or discriminability is achieved, because it shows that differentiation of the items does not begin until there is an opportunity for right *responses* to be reinforced while wrong ones are not. Finally, as the generalization begins to decrease, the curve of right responses rises reciprocally, showing that the items are becoming differentiated. We can infer, then, from our results as well as from our original hypothesis, that *a learned list is a differentiated one;* that is, one in which items that were once functionally equivalent have become discriminable.

This emphasis on differentiation as a part of the learning process is not unfamiliar, for students of animal learning have recently emphasized the role of discrimination in learning. Muenzinger (**11**), for instance, thinks that learning includes two different characteristics of behavior; learning how to discriminate in a new situation, and increasing efficiency of performance in successive repetitions of a behavior unit. Most discrimination experiments with animals differ considerably from the verbal learning situation, since instead of the usual two units to be discriminated, we are dealing with a number of units, usually presented in a temporal series. There are accordingly multiple generalization tendencies, and as these are broken down through differential reinforcement, many items are differentiated from one another, yielding finally a whole differentiated series. The complexity of discrimination may then be rather different in these two situations. But the fundamental process of differential response is common to both. It can certainly be concluded that the correlation between difficulty of discrimination of the items and rate of learning is a fact.

It is of interest to consider how widely the above principle can be related to other facts of memorization of single lists. Von Restorff (**13**) has made the only other study of intraserial similarity. She compared the retention-value of items surrounded by list-members of a different type with that of items surrounded by members homogeneous with them. Her major finding, that an 'isolated' item is recalled better than one in a homogeneous setting, is in line with the present hypothesis, since these conditions would represent, respectively, low and high generalization with the rest of the list, and the condition of low generalization should of course give better recall. She also advanced the hypothesis that nonsense syllables are difficult to learn primarily because of their homogeneous character—because the standard series consists of elements all of the same kind. The writer is in general agreement with this suggestion, since a list of nonsense syllables would represent a condition of very high generalization. Not only are the syllables alike, to begin with, in being all composed of 3 letters in accordance with a single set of rules; but their relatively meaningless character itself implies that they are also relatively undifferentiated. As items in a nonsense list become differentiated during learning, they not only take on distinction and identifiability, but successful association with a response item may give the stimulus item a certain symbolic reference as well. We can assume that meaningful units are ones which have been differentiated in the course of past learning from an original mass of equivalents. Lack of meaning on the other hand implies lack of differentiating characteristics. In the light of the present hypothesis, the meaninglessness and the homogeneity of nonsense items are both aspects of the fact that they are undifferentiated, and it is for this reason that they are difficult to learn.

Other intraserial effects which have been ascribed to inhibitory factors may be relevant to the hypothesis as well. Woodworth (**14**, p. 231) has suggested that retroactive and proactive inhibition will explain the well-known fact that as lists are made longer they become disproportionately harder to learn. The writer has already shown how the present hypothesis applies to retroactive and proactive inhibition between lists (**4**). In connection with the length of list phenomenon, it can be pointed out that as the list is made longer, the number of generalizing tendencies increases approximately as the square of the number of items,[7] increasing the opportunities for blocking of right responses and necessitating a longer process of differential reinforcement.

[7] The formula for the number of potential generalizing tendencies within a list of paired items is $n(n-1)$.

Foucault (**2**) and Gibson and Raffel (**5**) have suggested retroactive and proactive inhibition as the explanation of certain serial position effects. The present hypothesis has not been elaborated with a view to explanation of serial position effects, but there are some indications that generalization may play a part in these effects. For one thing, neighboring positions in a list would undoubtedly provide, themselves, a basis for generalization. For another thing, the well-known serial position phenomena have always been demonstrated with homogeneous lists of poorly differentiated material—nonsense syllables, geometric forms, or digits. It may well be asked whether these effects would be so apparent in a list made up of highly differentiated items; and whether their absence might not indicate the importance of generalization as a determining factor.

Summary and Conclusions

1. Lists of forms paired with nonsense syllables have been shown to be harder to learn when the forms within the list have a relatively strong tendency to generalize with one another, as compared with a list whose stimulus-forms have low generalizing tendencies. Furthermore, high-generalization lists are not so well retained as low-generalization lists, even though the former have been given many more repetitions.

2. When a particular form-syllable pair is a member of a list containing forms which generalize with it to a high degree, more repetitions are required before the response is first given correctly than when no forms having a strong tendency to generalize with it are present in the list.

3. A greater number of errors of overt generalization occurs in a high-generalization list than in a low-generalization list, during both learning and recall. Moreover, such errors occur with a frequency proportional to the previously measured degree of generalization between the stimulus items concerned.

4. In a list of forms paired with syllables, the amount of reinforcement required to set up differentiation between a particular pair of forms increases with increasing strength of the generalization between them.

5. The curve of generalization during the learning of a list shows a rapid initial rise followed by a relatively slow fall. The curve of learning rises parallel with the curve of generalization during the first three tenths of the learning process, and the two are reciprocal during the last five tenths of the learning process.

6. These conclusions suggest that differentiation of items is an essential component of verbal learning, and that differential rein-

forcement of right and wrong (generalized) responses is the process whereby a differentiated series is achieved.

(Manuscript received July 27, 1941)

References

1. Beritoff, J. S., On the fundamental nervous processes in the cortex of the cerebral hemispheres. I. The principal stages of the development of the individual reflex; its generalization and differentiation, *Brain*, 1924, **47**, 109–148.
2. Foucault, M., Les inhibitions internes de fixation, *L'Ann. Psychol.*, 1928, **29**, 92–112.
3. Gibson, E. J., A systematic application of the concepts of generalization and differentiation to verbal learning, *Psychol. Rev.*, 1940, **47**, 196–229.
4. Gibson, E. J., Retroactive inhibition as a function of degree of generalization between tasks, *J. exp. Psychol.*, 1941, **28**, 93–115.
5. Gibson, J. J., and Raffel, G., A technique for investigating retroactive and other inhibitory effects in immediate memory, *J. gen. Psychol.*, 1936, **15**, 107–116.
6. Hovland, C. I., The generalization of conditioned responses. III. Extinction, spontaneous recovery, and disinhibition of conditioned and of generalized responses, *J. exp. Psychol.*, 1937, **21**, 47–62.
7. Hovland, C. I., The generalization of conditioned responses. IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses, *J. exp. Psychol.*, 1937, **21**, 271–276.
8. Hull, C. L., The problem of stimulus equivalence in behavior theory, *Psychol. Rev.*, 1939, **46**, 9–30.
9. Hunter, W. S., Learning: IV. Experimental studies of learning, Chap. 11 in *Hanbk. Gener. Exper. Psychol.* (C. Murchison, Ed.), Worcester, Clark Univ. Press, 1934, 497–570.
10. Melton, A. W., The end-spurt in memorization curves as an artifact of the averaging of individual curves, *Psychol. Monog.*, 1936, **47**, No. 212, 119–134.
11. Muenzinger, K. F., *Psychology*, World Press, Inc., Denver, 1940, pp. 343.
12. Pavlov, I. P., *Conditioned Reflexes*, trans. by G. V. Anrep, London, Oxford Univ. Press, 1927, pp. 430.
13. von Restorff, H., Analyse von Vorgängen im Spurenfeld. I. Über die Wirkung von Bereichsbildungen im Spurenfeld, *Psychol. Forsch.*, 1933, **18**, 299–342.
14. Woodworth, R. S., *Experimental Psychology*, Henry Holt and Co., New York, 1938, pp. 889.

A STUDY OF THE AFFECTIVE NATURE OF THE INTERPOLATED ACTIVITY AS A FACTOR IN PRODUCING DIFFERING RELATIVE AMOUNTS OF RETROACTIVE INHIBITION IN RECALL AND IN RECOGNITION *

BY T. E. McMULLIN

University of Pennsylvania

Introduction and Statement of the Problem

Concurrently with the development of experimental work in the field of retroactive inhibition, has grown the disposition to utilize the concept as a major factor in the explanation of forgetting. The theory is that forgetting is not a function of elapsed time, per se, but of what happens *in time*. Partially as a result of this tendency, experimental work in the field of retroactive inhibition has become relatively more important as an approach to the study of the larger field of retention. Much experimental work has been done with various types of original learning materials—pictures, prose, poetry, word lists, nonsense syllables, numbers, geometric figures, etc. have been used.[1] A similarly wide range of materials has been used to provide the interpolated activity. For the rest condition, various experimenters have employed reading from a newspaper, marking the three best jokes on a page, casual conversation, playing simple games, and for the longer periods, 'normal activity,' sleep, and hypnosis. It is agreed that, for short interpolated intervals, some type of non-strenuous mental activity should be provided—except in instances where sleep or hypnosis is used—in order to prevent review of the original learning materials.

The length of the interpolated time interval has been subject to considerable experimentation. Intervals from a few minutes to several days, or weeks, have been used. Obviously, when the interpolated interval lasts for days, the interpolated learning activity

* This study was carried out in the psychological laboratory at the University of Kentucky, under the supervision of Dr. M. M. White.

[1] An excellent review of the literature on retroactive inhibition up to and including the year 1934, is given by S. H. Britt, in Retroactive inhibition: A review of the literature. *Psychological Bulletin*, 1935, **32**, 381–440.

cannot be kept up continuously, but is placed at some definite point in the total interval.

The final step in the retroactive inhibition set-up is the measurement of the retention of the original learning material. Of the three most commonly used general methods of measuring retention—recognition, recall, and relearning—the latter two have been used far more frequently than has recognition. One reason for this is that the phenomenon of retroactive inhibition appears to be more difficult to obtain with the use of recognition memory, and when obtained, is almost invariably in smaller amount. Until a few years ago, it was generally accepted that retroactive inhibition does not appear when recognition is used as the method of measuring retention. Recently, however, this conclusion began to be questioned and the problem was made the subject of more thorough experimental investigation. The resulting evidence, notably that obtained by McKinney (11) and Gibson (7), indicates that retroactive inhibition can be obtained when the recognition method of measuring retention is used and that the amount is almost invariably relatively smaller than the amount obtained when the recall method of measurement is used. There appears the possibility of a series of investigations, the results of which might be expected to show whether the phenomenon is subject to the same types of conditions in the two methods of measurement, and whether the same factors operate to produce it in recall and in recognition.

Two specific factors which may be studied are (1) the subject's attitude toward the interpolated activity, and (2) his emotional reactions during the interpolated learning activity. Bartlett (1) holds that in reproduction (recall) memory and in recognition memory an important consideration is the persistence (to some degree at least) of an original psychological orientation, or attitude, from the time of the learning of an activity until the time of the measurement, or use of the activity. He emphasizes this especially in the case of recognition, and says, "The essential difference between recognizing and remembering lies, however, not in an increase of complexity in the latter, but in a genuine difference in the way in which the necessary setting or scheme comes into play." If the implication involved here is correct, then it appears that in a retroactive inhibition set-up, an interpolated activity which markedly affects the orientation or attitude of the subject, or which is accompanied by marked emotional changes, may be expected to produce relatively different decrements in the retention of the original material, depending upon whether retention is measured by recall or by recognition. The following experiment was planned to test this hypothesis.

Procedure

1. *General Plan.*—The general plan of the experiment provided for the partial learning of three sets of original material, followed respectively by three periods of interpolated activity of equal duration but of supposedly different degrees of pleasantness. The interpolated activities were ranked by the subject as (1) 'best liked,' (2) 'next best liked,' and (3) 'least liked.' The interpolated conditions were also ranked by the experimenter, on the basis of certain physiological changes, as (1) 'least emotional,' (2) 'next least emotional,' and (3) 'most emotional.' Following each interpolated period, a recall test and a recognition test on the original material were given. The results of the recall and recognition tests were examined to determine whether the 'affective' and 'emotional' qualities of the interpolated activities bore the same relationships to the amounts recalled and the amounts recognized.

2. *Learning Materials.*—The original learning materials consisted of three lists of fifteen six-letter, two-syllable, common nouns. The interpolated materials consisted of (1) a collection of short stories, (2) a list of fifteen five-letter familiar adjectives, and (3) a list of fifteen five-letter nonsense words—paralogs.

3. *Apparatus.*—On the wall, directly in front of the subject, and about on a level with his eyes when he was seated, was a wooden arm, on the end of which was secured a Ranschburg exposure apparatus. This arm was hinged to the wall in such a way that it could be moved to one side to facilitate changing of the materials to be used, and then could be swung back and secured in a position convenient for the subject to see the materials.

A blood pressure cuff was attached to the subject's left ankle and the subject rested his left leg on a low chair. With the subject's left arm lying in a comfortable position on the chair arm, Darrow equal pressure electrodes were attached to the forearm and palmar surfaces. The connections from the electrodes and the tubing from the blood pressure cuff passed through the wall into the experimenter's room where they were attached to a Darrow Behavior Polygraph. Also, in the experimenter's room was a metronome which was in circuit with the exposure apparatus and regulated the time of exposure of each item. In the wall between the two rooms was a small opening, covered by a fine mesh screen, through which E could observe the behavior of S.

4. *Presentations.*—Before presentation of the original learning material, the following directions were given to S, orally: "We shall now have presented, one time, a list of words. While they are being presented you should concentrate on them and try to learn and remember (not necessarily in order) as many of them as you can, from the *one* presentation." The list was then presented one time at the rate of one item each two seconds.

When the interpolated activity was reading, the subject was instructed to select one of the short stories provided and begin reading silently. He was informed that there would be no test on the material read. When the interpolated activity was the learning of adjectives or paralogs, S was instructed to learn (or attempt to learn) the list by the anticipation method, saying aloud all items which he could anticipate. The items of the appropriate interpolated learning material were then presented at the rate of one each two seconds, with a rest interval of ten seconds at the end of the list each time it was presented. The subject was not informed whether there would be any test on these materials. Actually, no test was given—E simply observed that S kept himself occupied with the task.

The total working time for each subject was forty-three and one-half minutes. Each of the three periods of the sitting required fourteen and one-half minutes, divided as follows:

½ minute for the presentation of the original material.
10 minutes for the interpolated activity (including approximately 1½ minutes used for changing materials and giving instructions).
2 minutes for the recall test.
2 minutes for the recognition test.

Upon completion of each period, E informed S "You will not be tested again on any of the materials which you have had up to this point." The order of presentation of the lists and the order of the interpolated activities were rotated in such a way as to give equal average position to the variables and to provide for all possible order combinations.

5. *Measurement.*—The following instructions for the recall test were given orally: "Write down on this blank paper as many as you can remember of the words in the list which was presented to you *one time* just prior to the last activity." The recognition test utilized a slip of paper,

on which were mimeographed forty-five six-letter, two-syllable, common nouns. The fifteen words comprising the original material were included in this list and were scattered at random throughout the list. At the top were these directions: "In the list below you are to UNDERLINE each word which you recognize as having been presented to you a few minutes ago." A separate recognition test was necessary for each of the three lists of original material

The formula for scoring the recall test was *Number Right.* The formula for scoring the recognition test was *Number Right minus one-half Number Wrong.*

6. *Determination of the 'Affective' Value and 'Emotional' Value for the Interpolated Activities.*—Upon completion of the sitting, the subject was asked to rank the three interpolated activities in the order in which he 'liked' them. This ranking was taken as an indicator of his *attitude* toward the activities.

The polygraphic record of certain physiological changes in the subject, during a four-minute interval near the middle of each of the interpolated periods, was examined for evidences of changes indicative of emotion: (1) fluctuations in blood pressure, (2) increases in pulse rate, and (3) PGR phenomena. On the basis of changes, the three interpolated periods were ranked as (1) 'least emotional,' (2) 'next least emotional,' and (3) 'most emotional.'

7. *Subjects.*—Data were obtained on the performances of fifty-four women and eighteen men. All were university summer session students except one, who was, in all other respects, comparable to the students. Most of them had not had previous experience as subjects in psychological experiments.

RESULTS

The percents of retroactive inhibition produced by each of the work conditions both in recall and in recognition were detected and compared for each subject. In a small number of cases the score following the rest condition was zero, and the score following one of the work conditions was greater than zero. In such cases, the usual procedure of determining the percent of retroactive inhibition would indicate a *minus infinity* amount for a given subject. In order to avoid this difficulty, the following technique was used for all individual subjects. The score following the work condition was subtracted from the score following the rest condition, and the sign preserved. This difference was then divided by the *larger* of the two scores and the result was multiplied by 100. This procedure does not differ from the usual method except in those cases where there exists what apparently is a negative amount of retroactive inhibition, or facilitation. In such cases, the amount of this apparently negative retroactive inhibition is limited to 100 percent. This range of possible percents of retroactive inhibition for individuals, then is from −100 percent to +100 percent. Using this procedure, the percent of retroactive inhibition was computed for each subject for each of the work conditions, both in recall and in recognition.

The percent of retroactive inhibition produced by the adjectives in recall, for each subject, was subtracted from the percent produced by the paralogs in recall—this difference was called *a.* The percent of retroactive inhibition produced by the adjectives in recognition was subtracted from the percent produced by the paralogs in recognition—this difference was called *b.* If *a* and *b* are equal, then, apparently the adjectives and the paralogs have the same relation-

ship to each other as factors in producing decrements in recall that they have as factors in producing decrements in recognition. If *a* is greater than *b*, then it would appear that the adjectives operate to produce relatively more retroactive inhibition in recognition, and the paralogs relatively less. If *b* is greater than *a*, the reverse would appear to be true. When the scores of the seventy-two subjects were treated in the manner described, it was found that

b was greater than *a* in the cases of 39 S's (54.17 percent).
a was greater than *b* in the cases of 32 S's (44.44 percent).
a and *b* were equal in the case of 1 S (1.39 percent).

$$\sigma_P = 5.865.$$[2]

These results indicate that, as between the adjectives and the paralogs, there is a tendency for the adjectives to be relatively more effective than the paralogs in producing retroactive inhibition in recall, and for the paralogs to be relatively more effective than the adjectives in producing retroactive inhibition in recognition. Statistically, the chances are approximately 80 in 100 that this finding would be repeated, under the same conditions.

The data were also treated in terms of average scores. The same procedure was used as was used for the individuals, except that, since the presence of zero scores presented no difficulty, the percents of retroactive inhibition were computed according to the generally used formula:

$$\frac{\text{Ave. Rest Score minus Ave. Work Score}}{\text{Ave. Rest Score}} \times 100.$$

TABLE 1

AMOUNTS OF RETROACTIVE INHIBITION PRODUCED BY ADJECTIVES AND BY PARALOGS IN RECALL AND IN RECOGNITION (72 SUBJECTS)

Condition	Recall		Recognition	
	Ave. No. Words Recalled	Amount of R. I.	Ave. No. Words Recognized	Amount of R. I.
R. Reading (Rest)	3.29		8.81	
A. Adjectives	2.62	20.46%	7.72	12.38%
P. Paralogs	2.83	13.92%	7.90	10.33%

a. Superiority of A over P as a factor in producing R. I. in recall 6.54
b. Superiority of A over P as a factor in producing R. I. in recognition 2.05

Table 1 shows the results of this treatment of the data. In terms of averages, it is noted that both in recall and in recognition the adjectives produced a greater amount of retroactive inhibition than did

[2] In computing σ_P, the 1.39 percent of the cases showing no difference was divided equally between the other two conditions.

the paralogs. This superiority of the adjectives, however, is relatively greater in recall than in recognition. This trend is in agreement with the trend shown when the subjects' performances were treated individually. The problem arises: Is this relative difference due to the nature of the interpolated materials, or to the subjects' affective and emotional states during the interpolated periods, or to both?

One approach to the problem is to classify the two work conditions, not as adjectives and paralogs, but as the 'better liked' and the 'less liked' of the two work conditions, and see whether this treatment of the data shows noticeable differences in the relative amounts of retroactive inhibition in recall and in recognition produced by each of these two classifications. Table 2 shows these results.

TABLE 2

AMOUNTS OF RETROACTIVE INHIBITION PRODUCED BY THE 'BETTER LIKED' AND BY THE 'LESS LIKED' OF TWO INTERPOLATED ACTIVITIES IN RECALL AND IN RECOGNITION (65 SUBJECTS[3])

Condition	Recall		Recognition	
	Ave. No. Words Recalled	Amount of R. I.	Ave. No. Words Recognized	Amount of R. I.
1. Reading (Rest)	3.37		9.02	
2. 'Better liked' of the two work conditions	2.61	22.60%	7.89	12.53%
3. 'Less liked' of the two work conditions	2.92	13.24%	7.92	12.19%

a. Superiority of 2 over 3 as a factor in producing R. I. in recall 9.36
b. Superiority of 2 over 3 as a factor in producing R. I. in recognition ..... .34

The results indicate that the 'better liked' of the two work conditions tends to be more effective in producing retroaction than the 'less liked,' in recall, while in recognition, the two conditions operate with approximately equal effectiveness.

Table 3 shows the results when all three of the interpolated activities were classified on the basis of 'best liked,' 'next best liked,' and 'least liked' of the conditions. These results indicate that in recall, the interpolation of a somewhat less pleasant activity produces a greater decrement in memory than does the interpolation of a still less pleasant activity, whereas, in recognition, increasing the degree of unpleasantness tends to increase the amount of the decrement.

In Table 4 are shown the results when the two work conditions were classified on the basis of evidence of certain physiological phe-

[3] In the cases of seven subjects, the reports concerning their attitudes toward the three interpolated activities were not decisive, or were lacking; hence, only 65 subjects were considered in this classification.

TABLE 3

AMOUNTS OF DECREMENT PRODUCED BY THE 'NEXT BEST LIKED' AND BY THE 'LEAST LIKED' OF THREE INTERPOLATED ACTIVITIES IN RECALL AND IN RECOGNITION (65 SUBJECTS)

Condition	Recall		Recognition	
	Ave. No. Words Recalled	Amount of Decrement	Ave. No. Words Recognized	Amount of Decrement
1. 'Best liked' of the three activities	3.45		8.68	
2. 'Next best liked' of the three activities	2.53	26.56%	8.23	5.23%
3. 'Least liked' of the three activities	2.92	15.18%	7.92	8.77%

a. Superiority of 2 over 3 as a factor in producing a decrement from 1, in recall 11.38
b. Superiority of 3 over 2 as a factor in producing a decrement from 1, in recognition 3.54

nomena indicative of emotion, as determined by examination of the polygraphic records. The results of this approach indicate that an increase in the 'emotional' phenomena during the interpolated work condition is accompanied by an increase in the memory decrement produced, and that the relative increase in the memory decrement is greater in recognition than in recall.

TABLE 4

AMOUNTS OF RETROACTIVE INHIBITION PRODUCED BY THE 'LESS EMOTIONAL' AND BY THE 'MORE EMOTIONAL' OF TWO INTERPOLATED WORK CONDITIONS IN RECALL AND IN RECOGNITION (68 SUBJECTS[4])

Condition	Recall		Recognition	
	Ave. No. Words Recalled	Amount of R. I.	Ave. No. Words Recognized	Amount of R. I.
1. Reading (Rest)	3.28		8.75	
2. 'Less emotional' of the two work conditions	2.74	16.59%	8.03	8.24%
3. 'More emotional' of the two work conditions	2.63	19.96%	7.52	14.03%

a. Superiority of 3 over 2 as a factor in producing R. I. in recall 3.37
b. Superiority of 3 over 2 as a factor in producing R. I. in recognition 5.79

The results obtained when all three interpolated conditions were classified as 'least emotional,' 'next least emotional,' and 'most emotional,' in terms of physiological phenomena indicative of emotion, are shown in Table 5. These results indicate that the interpolation of a slightly emotional activity produces a greater decrement

[4] In the cases of four subjects, the data obtained on the polygraphic records were not considered sufficient to justify a ranking of the interpolated periods; hence, only 68 subjects were considered in this classification.

than does a still more emotional activity, in recall; whereas, in recognition, the most emotional activity produces the greatest decrement.

Another approach which can be made to the question of the relative effects of different interpolated activities upon recall and recognition is to note changes in absolute differences between the recall and recognition scores. Since recognition scores are almost invariably larger than recall scores, then, any interpolated activity which tends to reduce this difference is producing a relatively greater decrement in recognition. Any interpolated activity which tends to increase this difference is producing a relatively greater decrement in

TABLE 5

AMOUNTS OF DECREMENT PRODUCED BY THE 'NEXT LEAST EMOTIONAL' AND BY THE 'MOST EMOTIONAL' OF THREE INTERPOLATED CONDITIONS IN RECALL AND IN RECOGNITION (68 SUBJECTS)

Condition	Recall		Recognition	
	Ave. No. Words Recalled	Amount of Decrement	Ave. No. Words Recognized	Amount of Decrement
1. 'Least emotional' of the three activities	3.10		8.26	
2. 'Next least emotional' of the three activities	2.72	12.32%	8.46	−2.40%
3. 'Most emotional' of the three activities	2.82	9.24%	7.59	8.10%

a. Superiority of 2 over 3 as a factor in producing a decrement from 1, in recall..... 3.08
b. Superiority of 3 over 2 as a factor in producing a decrement from 1, in recognition.... 10.50

recall. Table 6 shows these average differences between the recall and recognition scores following the various interpolated conditions, when these conditions are classified according to the ways used in Tables 1, 2, 3, 4, and 5.

Assuming that conditions 1, 2, and 3, in Table 6, become less pleasant in that order,[5] it is noted that there is a tendency for the differences between the recall and the recognition scores to decrease toward the less pleasant end of the scale. This is especially true as between conditions 2 and 3, since by each of the five classifications, the average absolute difference between the recall and the recognition scores is less following condition 3 than following condition 2. This would seem to indicate that the increased degree of unpleasantness from condition 2 to condition 3 (by each classification) produces

[5] According to the averages of the subjects' rankings, reading was the 'best liked' activity and learning the paralogs was the 'least liked' activity. According to the average rankings made on the basis of the polygraphic records, reading was the 'least emotional' activity and learning the paralogs was the 'most emotional' activity. In the latter case, however, learning the paralogs was only slightly 'more emotional' than learning the adjectives.

TABLE 6

AVERAGE ABSOLUTE DIFFERENCES BETWEEN RECALL AND RECOGNITION SCORES FOLLOWING VARIOUS INTERPOLATED CONDITIONS
(Read Across)

Condition 1	(A) Ave. Diff. Between Recall and Recognition Scores	Condition 2	(B) Ave. Diff. Between Recall and Recognition Scores	Condition 3	(C) Ave. Diff. Between Recall and Recognition Scores	(D) B minus C	(E) σ Diff. (B minus C)	(F) Chances in 100 that B is greater than C
(*Cf.* Table 1) Reading	5.51	Adjectives	5.10	Paralogs	5.06	.04	.369	54
(*Cf.* Table 2) Reading	5.66	'Better liked' of two work activities	5.28	'Less liked' of two work activities	5.00	.28	.394	76
(*Cf.* Table 3) 'Best liked' of all three activities	5.24	'Next best liked' of all three activities	5.70	'Least liked' of all three activities	5.00	.70	.369	97
(*Cf.* Table 4) Reading	5.47	'Less emotional' of two work conditions	5.29	'More emotional' of two work conditions	4.90	.39	.370	86
(*Cf.* Table 5) 'Least emotional' of all three conditions	5.15	'Next least emotional' of all three conditions	5.74	'Most emotional' of all three conditions	4.77	.97	.411	99

TABLE 7

RATIO OF RECALL SCORES TO RECOGNITION SCORES, FOLLOWING VARIOUS INTERPOLATED CONDITIONS
(Read Across)

Condition 1	Ratio	Condition 2	Ratio	Condition 3	Ratio
(*Cf.* Table 1) Reading	2.68	Adjectives	2.95	Paralogs	2.79
(*Cf.* Table 2) Reading	2.68	'Better liked' of two work activities	3.03	'Less liked' of two work activities	2.71
(*Cf.* Table 3) 'Best liked' of all three activities	2.52	'Next best liked' of all three activities	3.25	'Least liked' of all three activities	2.71
(*Cf.* Table 4) Reading	2.67	'Less emotional' of two work conditions	2.94	'More emotional' of two work conditions	2.87
(*Cf.* Table 5) 'Least emotional' of all three conditions	2.66	'Next least emotional' of all three conditions	3.11	'Most emotional' of all three conditions	2.69

a relatively greater decrement in the recognition scores than in the recall scores. In none of the cases is the difference between these average differences equal to three times the standard error of the difference.

Another consideration is the change in the ratio of the recall scores to the recognition scores following the various interpolated conditions. The ratios are shown in Table 7. The amounts in the table show the number of times greater the average recognition score is than the average recall score. It is seen from these amounts that in all five of the methods of classifying the interpolated activities, or

TABLE 8

DATA ON THE PERFORMANCES OF THE TWENTY-TWO SUBJECTS WHOSE REPORTS AS TO THEIR ATTITUDES TOWARD THE THREE INTERPOLATED ACTIVITIES SHOWED EXACT RANK-ORDER AGREEMENT WITH THEIR POLYGRAPHIC RECORDS

Condition	Ave. No. Words Recalled	Ave. No. Words Recognized	Ave. Diff. Between Recall and Recognition Scores	Ratio of Recall to Recognition Scores
1. Reading	3.09	9.11	6.02	2.95
2. Adjectives	2.68	8.48	5.80	3.16
3. Paralogs	2.91	7.84	4.93	2.70
1. Reading	3.09	9.11	6.02	2.95
2. 'Better liked and less emotional' of two work conditions	2.73	8.91	6.18	3.27
3. 'Less liked and more emotional' of two work conditions	2.86	7.41	4.55	2.59
1. 'Best liked and least emotional' of all three conditions	3.36	9.16	5.80	2.72
2. 'Next best liked and next least emotional' of all three conditions	2.45	8.86	6.41	3.61
3. 'Least liked and most emotional' of all three conditions	2.86	7.41	4.55	2.59

conditions, there is an almost uniform pattern of change in the recall-recognition ratio as one considers the three conditions in the order of decreasing pleasantness. Without exception, the ratio increases following the intermediate condition and decreases following the least pleasant condition to a level near that following the most pleasant condition. This approach is simply another method of presenting the general trend shown in the data from this experiment, viz., the tendency for slight increases in the unpleasantness of the interpolated activity to produce a relatively greater decrement in recall memory, and for greater increases in unpleasantness to produce a relatively greater decrement in recognition memory.

It is of interest to note the performances of those subjects whose ranking of the three interpolated activities, in terms of their attitudes toward them, showed exact agreement with the ranking of the in-

terpolated conditions, made on the basis of the polygraphic records. In twenty-two cases there was such agreement. The data for these subjects are shown, in terms of average scores, in Table 8. These averages show the same trends which are shown in the group as a whole. From these data it would appear that when increasing dislike for the interpolated activity is reinforced by increasing physiological evidences of emotion, there is a tendency for the increased unpleasantness (beyond a certain point) to produce a relatively greater decrement in recognition than in recall.

Additional evidence of the difference between recall and recognition is indicated in the manner in which they operated in the different time intervals in the work period of the experiment. A recall test and a recognition test were given to each subject at the end of each third of the total working time. The averages of the recall scores, by periods, show a downward trend, while the averages of the recognition scores show an upward trend. These data would indicate

(72 Subjects)	Period		
	I	II	III
Ave. no. words recalled	3.04	2.97	2.74
Ave. no. words recognized	7.90	8.13	8.40
Ave. diff. between recall scores and recognition scores	4.85	5.16	5.66
Ratio of ave. recall score to ave. recognition score	2.60	2.74	3.07

that recall is more susceptible to decremental effect from whatever degree of fatigue was produced in the total working period, than is recognition. They would also indicate that recognition is more susceptible to decremental effect from whatever degree of 'newness,' uncertainty, or apprehension, that may have been experienced by the subjects in the earlier stages of the total working period, and which may have subsided as the total period progressed. In terms of the subjects' reports, the interpolated activity in the first period was 'least liked,' on the basis of average rankings, and the interpolated activity in the third period was 'best liked.' In terms of the polygraphic records, the interpolated condition in the second period was 'least emotional,' and the interpolated conditions in the first and third periods were approximately equal.

Conclusions

There are three major theories of retroactive inhibition. One is the perseveration theory, which holds that retroactive inhibition is

produced when the interpolated activity interferes with the perseveration activity, or 'setting-in' process of the original learning. The implication is that the immediacy and the intensity of the interpolated activity are determining factors in producing the decrement.

Another theory is the transfer theory, which holds that certain learning elements common to both the original activity and the interpolated activity operate to confuse the subject in the test of retention. The implication is that the degree of similarity of the original and the interpolated activities is a determining factor in producing the decrement.

The third theory is the transfer theory and disruption hypothesis, which is similar to the transfer theory, with the additional hypothesis that certain learning elements in the original activity transfer to the interpolated activity, thereby disrupting the organization of the former.

The present experiment is not considered crucial in regard to the theories of retroactive inhibition, but the trends shown in the data suggest the possibility that any theoretical explanation of the phenomenon should be made only in terms of the method used for measuring retention. It is conceivable that the perseveration theory may be most applicable when the recognition method of measuring retention is used; the transfer theory may be the best explanation of the phenomenon when obtained in terms of recall memory; and the transfer theory and disruption hypothesis may best fit the facts when the relearning method is employed.

The trends in the data in this experiment are summarized in the following statements.

1. When original learning material consisted of a list of familiar nouns, an interpolated list of familiar adjectives produced greater amounts of retroactive inhibition both in recall and in recognition than did a list of paralogs. The relative superiority of the list of adjectives, however, was greater in recall than in recognition.

2. When the two work conditions were classified as 'better liked' and 'less liked,' the 'better liked' activity produced a greater amount of retroactive inhibition in recall than did the 'less liked' activity. In recognition, the amounts produced by the two classifications were approximately equal.

3. When all three interpolated activities (including reading) were classified as 'best liked,' 'next best liked,' and 'least liked', the 'next best liked' activity produced a greater decrement than did the 'least liked' activity, in recall. The reverse was true in recognition.

4. When the two work conditions were classified as 'less emotional' and 'more emotional,' on the basis of certain physiological phenomena, the 'more emotional' condition produced greater amounts

of retroactive inhibition both in recall and in recognition, but its superiority as a factor in producing retroaction was slightly relatively greater in recognition.

5. When all three of the interpolated conditions were classified as 'least emotional,' 'next least emotional,' and 'most emotional,' on the basis of certain physiological phenomena, the 'next least emotional' condition produced a greater decrement than did the 'most emotional' condition, in recall. The reverse was true in the case of recognition.

6. In considering these five pairs of interpolated conditions: (1) adjectives and paralogs, (2) 'better liked' and 'less liked' of the work conditions, (3) 'next best liked' and 'least liked' of all three conditions, (4) 'less emotional' and 'more emotional' of the work conditions, and (5) 'next least emotional' and 'most emotional' of all three conditions, the average absolute difference between the recall and the recognition scores was less following the latter named condition in every case.

7. In considering these five pairs of interpolated conditions: (1) adjectives and paralogs, (2) 'better liked' and 'less liked' of the work conditions, (3) 'next best liked' and 'least liked' of all three conditions, (4) 'less emotional' and 'more emotional' of the work conditions, and (5) 'next least emotional' and 'most emotional' of all three conditions, the ratio of the average of the recall scores to the average of the recognition scores was lower following the latter named condition in every case.

8. Of two interpolated work conditions, one 'more pleasant-less emotional,' the other 'less pleasant-more emotional,' either may be superior as a factor in producing retroactive inhibition, but the chances of the latter's being superior are greater in recognition than in recall.

(Manuscript received June 24, 1941)

References

1. Bartlett, F. C., *Remembering*, London, 1932, Cambridge University Press.
2. Bills, A. G., *General Experimental Psychology*, New York, 1934, Longmans, Green and Co.
3. Britt, S. H., Retroactive inhibition: A review of the literature, *Psychol. Bull.*, 1935, 32, 381–440.
4. Britt, S. H., Theories of retroactive inhibition, *Psychol. Rev.*, 1936, **43**, 207–216.
5. Crosland, H. R., A qualitative analysis of the process of forgetting, *Psychol. Monog.*, 1921, **29**, No. 130.
6. DeCamp, J. E., A study of retroactive inhibition, *Psychol. Monog.*, 1915, **19**, No. 84.
7. Gibson, J. J., Retroaction and the method of recognition, *J. gen. Psychol.*, 1934, **10**, 234–236.
8. McGeoch, J. A., On the term 'retroactive inhibition,' *Amer. J. Psychol.*, 1930, **42**, 455–457.
9. McGeoch, J. A., Forgetting and the law of disuse, *Psychol. Rev.*, 1932, **39**, 352–370.
10. McGeoch, J. A., and McKinney, Fred, Retroactive inhibition in the learning of poetry, *Amer. J. Psychol.*, 1934, **46**, 19–33.

11. McKinney, F., Retroactive inhibition and recognition memory, *Psychol. Bull.*, 1933, **30**, 699–700.
12. Myers, G. C., A comparative study of recognition and recall, *Psychol. Rev.*, 1914, **21**, 442–456.
13. Twining, P. E., The relative importance of intervening activity and lapse of time in the production of forgetting, *J. exper. Psychol.*, 1940, **26**, 483–501.
14. Webb, L. W., Transfer of training and retroaction. A comparative study, *Psychol. Monog.*, 1917, **24**, No. 104.
15. Zangwill, O. L., A study of the significance of attitude in recognition, *Brit. J. Psychol.*, 1937–38, **28**, 12–17.
16. Zangwill, O. L., The problem of retroactive inhibition in relation to recognition, *Brit. J. Psychol.*, 1937–38, **28**, 229–247.

MEMORY TRACE FOR COLOR [1]

BY NELSON G. HANAWALT AND BARBARA E. POST

New Jersey College for Women, Rutgers University

The present study is an attempt to test Koffka's hypothesis that the memory trace for color progresses toward a more saturated color in retention, in keeping with the Gestalt hypothesis of progressive change in the trace toward a 'better' figure in memory for designs. If a saturated color is a 'better' color, it would appear that the hypothesis of progressive change in the trace could be more readily tested here since the concept of 'better' figure is a very elusive one.

Koffka (**6**, p. 497 f.) reports that Wulf carried on some experiments with colors of different saturations. However, although the first results were significant enough, he did not have time to develop them systematically so they were never published. He refers to the work of Katz as additional evidence. Katz finds that if a subject is asked to select from a series of tints and colors, the one that is equal to the blue of his friend's eyes, the black of his own hat, or the red of his lips, he will as a rule select too saturated a color.

Since Hanawalt (**3**) finds little evidence for progressive change in the memory trace for designs during retention it was considered quite worth while to test this further hypothesis in the realm of color. It appeared especially desirable to make this test since the hypothesis is based upon very thin evidence. The color of a friend's eye is a very changeable standard and varies greatly with many environmental conditions. The color of one's black hat is a little more stable, but the color of one's lips is again likely to vary a great deal with physiological state and environment, not to mention the complications which are likely if the lips are those of a female!

Experiment I. Preliminary Experiment

Following Katz (**5**) we set up a recognition experiment. Red, green, and blue were used. A recognition test card consisting of nine one inch squares of color of varying amounts of saturation was constructed for each color. These squares were pasted on the card in random order.

[1] The authors are indebted to Miss Katherine Schick and Miss Doris Loncore for assistance in taking data for the preliminary experiment, to Miss Helen Losee for assistance in the second experiment, and to Miss Frances Potter for assistance in the fourth experiment. We are indebted to Professor Harry Helson of Bryn Mawr for the suggestion of using the color wheel. We gratefully acknowledge the assistance of the National Youth Administration in furnishing subjects for Experiment III.

Unfortunately paper of varying degrees of saturation in equal steps and limited to the variable of saturation alone was not available. This is a defect which can be remedied, and perhaps it would be worth while to repeat this experiment with this improvement. We shall mention the results briefly for they are interesting in light of the later experiments.

From each of the three recognition test cards we selected three colors of different saturation for stimulus colors—one from each end of each series and one from the center of the series. These stimulus cards were constructed using the same paper for both the stimulus and the background as we had used in constructing the recognition cards. Thus we had three recognition cards and nine stimulus cards.

Method.—The O's were stimulated and tested individually. The illumination was daylight. There were two test periods—immediate and after one week. Different groups of subjects were used for each degree of saturation and for each time interval. Thus there were six groups of subjects, sixteen to the group, making a total of 96 O's. The O's were all female and students at New Jersey College for Women. Each stimulus color was presented for fifteen seconds.

Results.—A summary of the results of this experiment are presented in Table I. The most interesting fact revealed in the results is the difference in responses for the two time intervals. Consequently we have lumped together the results for the different degrees of saturation and for the different colors. As will be noted in Table I, the results of the immediate group are in favor of Koffka's theory: thirty-four percent of the responses are in the direction of a more

TABLE I

Summary of Results of Preliminary Experiment

Time of Recog.	Stimulus Selected %	Error More Saturated %	Error Less Saturated %	N. of Responses
Immediate	54	34	12	144
After 1 wk.	30	35	35	144

saturated color while only twelve percent are toward a less saturated color. However the results from the delayed recognition group are certainly not what one would expect from Koffka's theory. Here there is no increase in the percent of more saturated responses but a great increase in the percent of less saturated responses. This is just the opposite from what we should expect from the theory. It is as if all the increase in error exhibited in delayed recognition went in the direction of less saturation. This doesn't seem to make much sense. The decrease in the percent of correct selections is however in the direction we should expect from real forgetting.

If one were to consider the results of each time period separately a hypothesis could be projected. The results in the immediate recognition group do suggest Koffka's hypothesis that the trace gravitates toward a more saturated color. The results of the delayed recognition group suggest a chance or non-directional hypothesis. Considering the results as a whole they suggest further experimenting.

Experiment II. Modified Method of Limits

Since the procedure in Experiment I failed to yield any conclusive results we decided to improve our conditions and refine our method. We are convinced that the recognition method is a good one providing adequate colored papers are available. The use of the color wheel, as suggested by Professor Harry Helson, proved a very satisfactory means of varying the saturation.

The plan of this experiment is to present the subject with a color of known saturation on the color wheel and then to approach this standard from either direction until the O signals that the saturation now equals the standard. To facilitate this procedure the Marietta wheel was used in order that the saturation could be varied without stopping the wheel. If the trace goes toward a more saturated color in retention we should expect that the judgments would be skewed accordingly. This would seem to be a fair test of the theory.

Apparatus.—A box screen 26 in square by 5 in deep was clamped on a table in front of O. O observed the wheel through a slit cut in the screen. A blind could be raised or lowered over this slit by a pulley arrangement. In order to cut out a view of the center and edges of the wheel, and to conceal E's operation of the saturation control, a second funnel-shaped screen was attached to the inside of the box screen over the slit with the small end pointing toward the wheel. The wheel was 4 in in front of the funnel screen. When the slit was open O observed a film color in the end of the funnel. The illumination was furnished by two daylight mazda bulbs (60 w.) placed in diagonally oposite corners of the box screen. O signaled his response to E by means of a signal light. A stop watch was used to time the length of exposure and the rate of change of saturation.

Subjects.—There were fourteen female subjects: seven of these were college juniors majoring in psychology, two were high school students, three were high school graduates (office workers), and two were college graduates (office workers). All were under twenty-five years of age.

Procedure.—The experiment was conducted in a dark room. The O was seated at the table 12 in in front of the box screen. Instructions to the subject were as follows:

"The cover over this slit will be raised for two seconds and you will see a color of a definite saturation in the end of the funnel. (At this point saturation was explained to the subjects.) At the end of two seconds the cover will be dropped. During this time the saturation will be changed, making it either more or less saturated. The cover will then be raised again, and you will see a color either more or less saturated than the standard which you saw during the two second interval. As you watch, the saturation of the color will be changed gradually so that the color approaches the standard color. As soon as you think the color is exactly the same as the first color which you observed, the standard, press the telegraph key. You will be given twenty ascending and twenty descending trials for each color, and before each trial you will see the standard color for two seconds."

Several trials were given before the experiment proper began to make sure that the subjects understood the directions and to acquaint them with the situation. The standard saturation was 80 percent color and 20 percent dark gray. The colors used were red, green, and blue. The changes in saturation ranged from 10 to 20 percent on each side of the standard in order that the subject would not make a judgment on a time basis. As soon as the subject pressed the key indicating that the color had now approached the standard a recording was taken and a new standard presented to the subject. A one minute interval was allowed between trials.

Results.—Table II gives a summary presentation of the results of this experiment. In Table III we have listed the average ascending and descending scores for each of the fourteen subjects to show the

performance of each observer. We have listed the three colors separately for there was some slight difference in performance on the different hues. In Table II each ascending average and each descending average is based upon the average of the average score for

TABLE II

AVERAGE POINT AT WHICH THE SUBJECTS JUDGED THE SATURATION OF THE HUE TO EQUAL THE STANDARD

The standard is 20 on the scale or 80 percent hue and 20 percent dark gray.*

Hue	Av. Ascending	Av. Descending	Av. Ascending-Descending
Red	22.13	18.60	20.37
Green	22.74	15.60	19.17
Blue	22.78	16.86	19.82
			Av. 19.79

* The scale on the wheel reads in terms of gray hence the lower the number the more saturated the mixture. Ascending and descending in the table, however, are in terms of saturation.

each subject's twenty judgments. In Table III each figure represents an average of forty judgments per subject.

When the comparison judgment is moving from a less saturated hue to a more saturated hue (see Table II) there is a definite tendency for the O's to show a predominance of less saturated responses. When the comparison stimulus is moving from a more saturated stimulus toward the standard there is a definite tendency for the O's to show a predominance of more saturated responses. This is consistently shown for all three hues and is true for practically all the individual averages. When the ascending and descending judgments are aver-

TABLE III

THE AVERAGE ASCENDING-DESCENDING JUDGMENTS FOR THE FOURTEEN SUBJECTS FOR THE THREE HUES

Red	Green	Blue
19.425	18.775	18.450
19.325	19.625	18.225
21.150	18.500	20.825
21.425	18.450	21.700
20.425	20.070	19.170
18.650	19.470	19.375
18.125	18.625	19.200
22.050	19.000	19.675
21.530	18.750	18.325
20.725	21.700	19.825
20.200	19.725	22.700
21.400	18.925	20.425
20.225	18.275	20.900
20.525	18.500	18.675

aged there is no reliable deviation from the standard. The standard in each case was 80 percent hue and 20 percent dark gray. It will be noted that green and blue show more variation than red, but the difference is largely on the descending series. This is probably due to the effect of the dark gray upon the saturation. Certainly the consistency in the ascending series is due to a great extent to the qualitative change which the dark gray introduces. This was borne out by the introspections of the subjects.

There is little evidence in these results that the trace goes toward a more saturated color in retention. Apparently there is a fairly wide area of indifference or equality. When the subject first reaches this area he responds 'equal,' hence his average falls short of the standard in both the ascending and the descending directions. When the ascending and descending judgments are averaged the differences from the standard cancel each other. The results for green are most favorable to Koffka's theory; red gives results in the opposite direction and blue only very slightly in the direction of the theory. When averaged there is only .2 percent divergence from the standard. Neither do the results from the individual subjects show any trend (see Table III). Only four of the fourteen subjects show a consistent trend for all three colors. It would appear that the results could be best accounted for on the basis of chance variation.

Experiment III. Modified Method of Constant Stimuli

Experiment II seemed to show that there was no trend toward a more saturated color in the trace, but it left us unsatisfied for the dark gray which we used for varying the saturation also varied the brightness and introduced qualitative changes. We decided to set up an experiment using the standard wheel so that we could use better saturated paper and control the factor of brightness. Hence the apparatus for this experiment is identical with that of Experiment II save for the change of color wheels and the discarding of the signal light.

The experiment is set up in such a way that every judgment is a test of the theory. All of the judgments are made around the threshold. There is no standard in the usual sense of the word, but the stimuli vary on either side of 50 percent saturation. A sample of the settings will best describe this method.

Percent of hue: 50, 48, 50, 50, 52, 50, 50, 48, 50, 50, 52, etc.

The O starts off with the observation of a saturation of 50 percent hue, then he is shown one of 48 percent and asked to judge whether this one is more or less saturated than the former. This 48 percent

is then the standard for the next comparison which is 50 percent, and this is the standard for the next judgment which is also 50 percent. Thus the stimulus is judged in comparison with the preceding stimulus, and that one was judged on the basis of the one that preceded it. Now if the hypothesis set forth by Koffka is correct we should expect the following results:

1. When the comparison stimulus is less saturated than the preceding stimulus there should be a predominance of correct responses because the color has become less saturated and the trace more saturated, thus creating a larger gap between the pair.

2. When the comparison stimulus has become more saturated than the preceding stimulus the correct and incorrect responses should approach a chance distribution for both the hue and the trace are moving toward more saturation.

3. When the comparison stimulus equals the preceding stimulus there should be a predominance of 'less' responses for the saturation of the stimulus has remained the same but the saturation of the trace has increased.

Subjects.—The results presented below are based upon extensive judgments of one subject. She was a very conscientious, reliable, and excellent subject. She worked quietly, patiently, and without asking any questions. Naturally she didn't know that every third stimulus was the same and doesn't know to this day. She knew nothing about the purpose of the experiment more than that we were interested in determining how small a discrimination she could make. She was being payed by the hour on a N.Y.A. project. We secured some results on a second subject whose results are not presented here because the school year ended before we were able to get anywhere near a complete series of results. Her attitude was not nearly as ideal but even so the results which we secured using her as a subject were essentially the same as those reported below.

Further Note on Procedure.—Since it was necessary to keep the subject from discovering that every third stimulus was the same as the one preceding, and since we were looking for very slight changes, we had to keep the judgments as near the threshold as possible. This was done by varying the size of the step on each side of the mid-point. When the subject was making good discriminations the step might be only 1 percent deviation on each side (but always the same percent on both sides) and when the subject was less efficient the step was increased. In this manner we kept the O in a position where she was always making some errors, but was making enough correct responses to assure us that she was making a discrimination. She was instructed that the discriminations would be very close and that she could make finer discriminations than she suspected that she could. She was told to make a decision each time of 'more' or 'less' and to guess if she was not sure.

Two different mid-points were used: one at 50 percent saturation and one at 75 percent saturation. Two time intervals between stimuli were used: one 28 sec. and one 58 sec. It was thought that comparing these two different time intervals would enlighten us concerning the progressive change if we should find one in the results. The exposure time was 2 sec., thus there was one trial every 30 sec. for the first time interval, and one every minute for the second time interval. As a general rule, no more than one hundred trials were run in one day under the minute cycle, and no more than fifty on any one color. Under the thirty second condition from one hundred and fifty to two hundred could be run at one sitting.

When all the trials were recorded, a tabulation was made showing the percent and number of correct and incorrect responses when the hue was more saturated than, less saturated than, or the same saturation as the one with which it was compared.

Results.—The results of this experiment are presented in Tables IV and V. In Table IV we have combined the judgments for the four hues, red, green, yellow, and blue, for there was no point in keeping them separated. Consequently all four hues are included in the figures in each line of Table IV. It will be seen immediately

TABLE IV

THE PERCENT OF CORRECT RESPONSES WHEN THE COMPARISON STIMULUS IS EITHER MORE OR LESS SATURATED THAN THE PRECEDING OR 'STANDARD' STIMULUS

Time Between Stimuli	When the 2nd S is More Saturated		When the 2nd S is Less Saturated	
	N.*	Percent Correct	N.	Percent Correct
28″	290	69	292	63
58″	291	67	290	67
Saturation				
50%	290	69	292	68
75%	291	69	290	62

* N. in each case stands for the number of chances to make a correct response under the prescribed conditions. The reciprocal of the percent correct will give the percent incorrect since the O responded either 'more' or 'less.'

that the results presented in this table do not bear out the expectation of Koffka's theory. In the column "When the 2nd S is more saturated" we should expect that there would be about an equal percent of correct and incorrect responses since the trace, presumably, would be going in the same direction as the comparison stimulus, and hence, make a comparison judgment very difficult. Actually

TABLE V

PERCENT OF 'MORE' AND 'LESS' JUDGMENTS WHEN THE TWO STIMULI OF THE COMPARISON PAIR ARE OF EQUAL SATURATION

Time Between Stimuli	Percent 'More' Judgments	Percent 'Less' Judgments	Number of Judgments
28″	50.55	49.45	287
58″	50	50	271

the performance is just as good under this condition as it is "When the 2nd S is less saturated" where we should expect a preponderance of correct responses over that revealed in the first case. In this latter case the discrimination should be easy for the comparison stimulus has decreased in saturation while the trace presumably has increased. The percent correct in each case reveals that we succeeded very well in keeping just above the threshold. According to the theory there should have been a greater difference between the first and the second

column under the 58″ interval than under the 28″ interval. Actually the difference is less, but the difference is an unreliable one. It will also be noted that the difference in saturation makes no difference in the results. This point should be borne in mind in connection with the results of Experiment IV.

Table V represents the results of the third test set-up in this experiment. Since the comparison stimuli are equal (a fact not suspected by our subject) this would seem to be a very good test of the theory. If the theory were upheld we should expect a preponderance of 'less' judgments, for the trace having become more saturated would make the second stimulus appear less saturated. Furthermore if the trace is progressing toward more saturation in time, we should expect more 'less' judgments for the longer time interval. Neither of these expectations is met. There is an almost perfect chance distribution of responses.

The results of Experiment III do not confirm the theory. The results fail to fulfill a single one of the three expectations we outlined at the beginning of the experiment. In fact if there is any trend in the results at all it would appear that it is in the opposite direction. We shall take up this point in our discussion at the end of Experiment IV.

Experiment IV. The Effect of Longer Intervals of Time on the Memory Trace for Color

In the preceding two experiments the longest time interval between the comparison stimuli was only fifty-eight seconds. We thought it would be very interesting to test longer periods of time. The idea of introducing equal stimuli into Experiment III furnished the spring board for the present set-up. The idea is to show the O's a hue of a given saturation on one day and then at a later interval show them the same saturation and have them make a judgment of more or less. Here again we should expect, according to Koffka's theory, that there would be a predominance of 'less' judgments since the trace in the interval would progress toward a more saturated hue, hence the judgment made after a considerable lapse of time would be 'less.'

The apparatus used in this experiment is identical with that used in Experiment III. The procedure however was greatly modified. In place of having one subject make a great number of judgments we had many subjects make a very few judgments. The apparatus was in the dark room in the laboratory. The laboratory was also in semi-darkness. The subjects (classes in psychology) filed past the apparatus and had a 2 sec. view of the hue. At the next meeting of the class forty-eight hours later they again filed past and had another 2 sec. view of the same stimulus. The wheel was not changed in the meantime so that there would be no question of its not being the same. As described in Experiment II the stimulus was a film color. After the second view of the stimulus the O moved on to another table and wrote on a signed slip of

paper either 'more saturated' or 'less saturated.' The judgment of 'same' was permitted but was discouraged as will be shown below. No subject gave more than one judgment for the same hue in order that previous experience with a hue would not be a factor in a specific judgment. However each subject, for the most part, gave one judgment for each of four hues—red, yellow, green, and blue. The subjects were all female college students between the ages of seventeen and twenty-two.

Instructions to Subjects.—The O's were told that this was an experiment in color discrimination in the dimension of saturation. Saturation was carefully explained to them. They were also told that the discriminations would be very close and that they were to say either 'more saturated' or 'less saturated' if they possibly could. They were informed that they could make much finer discriminations than they thought was possible. It was stressed several times that no conversation at all would be permitted during the experiment. It was explained that a stray suggestion might influence another's vote. They were asked not to discuss their judgment with other students. Of course they had no idea about the real purpose of the experiment. They cooperated very well in the experiment and showed a real interest in the results. They were very eager to know how they were doing.

Results.—The results of Experiment IV are summarized in Table VI. Again they are negative as far as Koffka's theory is concerned. Since the stimuli are identical for the comparison, one would expect that the trace, presumably having become more saturated, would give rise to a predominance of 'less saturated' responses. The reverse is true (see Table VI). The fact that there is no difference between

TABLE VI

RESPONSE TO THE SAME STIMULUS AFTER FORTY-EIGHT HOURS

Saturation	Responses			
	More	Less	Same	*N*
% of hue	%	%	%	
95	62	36	2	91
75	45	45	10	116
20	60	37	3	327

the responses 'more' and 'less' for 75 percent saturation is baffling. This is especially so in light of the fact that the responses are almost identical for 95 percent and 20 percent saturation. One might venture to say that in the case of the 95 percent saturation the hue is more saturated than the usual hue to which the subject is accustomed but it is not anything to get excited about. He remembers it as a good green, blue, etc. On seeing it the second time he is again impressed with the unusual hue and is more inclined to respond 'more saturated' because in the meantime his trace has gravitated toward the average hue of everyday life. Apparently the 75 percent saturation is about an average saturation as far as the subject's everyday experience is concerned. When he sees the color the second time he is at a loss as to whether or not it is more or less saturated. Consequently he responds on a chance basis or evades the issue by re-

sponding 'same' as 10 percent of the O's did for this series compared with only 2 and 3 percent for the other two series.

Before the 20 percent saturation series was run one of the authors reported the results of these experiments before the Philadelphia Experimental Psychologists, giving an explanation for the two series as reported above. The explanation seemed to be well taken and it was predicted that in a series of weak saturation one might expect the trend to reverse, *i.e.*, one might expect a predominance of 'less' responses. It will be noted from Table VI that this expectation was not fulfilled. From the introspections of the O's, it would seem that a different attitude prevailed which would account for the trend of the responses. Twenty percent saturation is a very weak one. The subjects had to strain to see a saturation at all. They remembered the hue as a very faint one. On seeing it the second time they were set for a very faint hue and were surprised to see as much saturation as actually appeared. Thus there resulted a predominance of 'more' responses.

The difficulty with an interpretation as outlined above is that it can be too readily adapted to any kind of results. Perhaps it would be better merely to state that we find a negative time error to exist under certain conditions in the dimension of saturation and decline to discuss the situation any further on the grounds of insufficient data. Certainly the experiment was not designed to study the time error. Now that we are reasonably sure of its presence it would be well worth the time and trouble to study it further.

We ran a few additional series with 20 percent saturation to test the effect of some of our conditions on the results. Perhaps the results would have been different had we not used film color. To test this we removed the funnel-shaped screen. Under this condition the O looked through the slit and could see the whole wheel. This did not make any difference in the results. The responses were still of the order of 60 percent 'more' and 40 percent 'less.' We also find that shortening the time interval between the two presentations of the stimuli doesn't change the results to any appreciable extent (a change from forty-eight hours to forty-five minutes). This seems to give added strength to the interpretation that the subject's attitude is responsible for the asymmetrical responses.

Discussion of Results.—The results of these experiments indicate quite clearly that Koffka's hypothesis of progressive change in the memory trace toward a more saturated color is invalid. The more refined the method of investigation grows, the less support we find for the hypothesis. In the light of the long history of the negative time error it is not surprising to find a slight trend in the opposite direction from that predicted in Koffka's theory. As suggested

above, our data are too limited in these experiments to speculate to any length on the factors responsible for the appearance of the negative time error in these results. However, it might be well to raise the question of our failure to find one in Experiment II—the modified method of limits. It appears quite likely that the greater qualitative change on the ascending series, due to the very dark gray used to vary the saturation, is responsible for this. The subjects' introspections bear out this contention. Also the greater consistency of the points of subjective equality for the various hues on the ascending series, compared with the descending series (see Table II), argues to this effect.

In the light of the results of this experiment we have to come to terms in some way with the results reported by Katz. Koffka reports that his interpretation is not strictly that reported by Katz. Katz (**5**, p. 163 f.) is speaking about the effect of language on memory color in connection with color constancy when he reports the results quoted by Koffka concerning the matching of eye color, hat shade, lip color, etc. with hues and colors. For instance he speaks of how prone we are to exaggerate in language our statement of the color of objects. Some examples given are 'as white as snow,' 'a jet-black negro,' 'a face white as chalk,' etc. Bartlett (**1**), Carmichael *et al.* (**2**), Hanawalt (**3**), Hanawalt and Demarest (**4**) and others have found that naming of perceptual experiences certainly is a factor in determining the reconstruction of designs from memory. We do not doubt that Katz would find the same factor operating in memory for colors. The results of Katz need not conflict with ours at all. Certainly it need not be concluded from his results that there is a progressive trend in the memory trace toward a more saturated color.

Several problems have been raised in this paper which have been only partially solved. What results might you expect using the recognition method, as we did in Experiment I, but with adequately controlled stimuli? The comparison of the results from the methods of recognition and successive comparison would be interesting in its own right, aside from the particular problem raised here. The effect of language on the memory trace, the problem with which Katz was concerned could be very well introduced into such an experiment. Another problem is that of the negative time error. We had not anticipated that it would cross our path when we designed these experiments. Very probably Koffka did not have it in mind either when he set forth his hypothesis.

Summary and Conclusions.—The series of experiments reported in this paper are designed to test Koffka's hypothesis that the memory trace for color progressively changes in retention in the direction of a more saturated color.

In Experiment I the recognition method is used. The results are inconclusive, due, we believe, to the inadequacy of the colored papers used for stimuli.

In Experiment II we use a modified Method of Limits. The Marietta color wheel, used for varying the saturation, was not completely satisfactory but the results were definitely negative, *i.e.*, there was no trend.

In Experiment III we use a modified Method of Constant Stimuli, with the usual type of color wheel. The conditions were quite satisfactory. Again the results are negative as concerns the theory. There is a suggestion of a negative time error.

In Experiment IV we use the comparison method with forty-eight hours elapsing between the presentation of the standard and the comparison stimuli—actually the two are identical. The results are again negative. There is a definite indication of a negative time error which is the opposite from what would be predicted from Koffka's hypothesis.

We conclude that Koffka's hypothesis concerning the progressive change of the memory trace for color, as measured by the comparison method, is invalid. Further work is indicated to test the hypothesis in recognition. Further research is also indicated for the study of the negative time error which appeared in connection with our last experiment.

(Manuscript received July 2, 1941)

References

1. Bartlett, F. C., *Remembering: a study in experimental and social psychology*, New York: Macmillan, 1932.
2. Carmichael, L., Hogan, H. P., and Walter, A. A., An experimental study of the effect of language on the reproduction of visually perceived forms, *J. exp. Psychol.*, 1932, **15**, 73–86.
3. Hanawalt, N. G., Memory trace for figures in recall and recognition, *Arch. Psychol.*, 1937, No. 216.
4. —— and Demarest, I. H., The effect of verbal suggestion in the recall period upon the reproduction of visually perceived forms, *J. exp. Psychol.*, 1939, **25**, 159–174.
5. Katz, D., *The world of colour* (Translated by MacLeod, R. B., and Fox, C. W.), London: Kegan Paul, Trench, Trubner, 1935.
6. Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace, 1935.

DEGREE OF EFFORT: I. AN INVESTIGATION OF A CONCEPT IN THE FIELD OF MOTIVATION

BY G. K. YACORZYNSKI

Northwestern University Medical School[1]

The concept of the degree of effort was formulated in an attempt to explain the atypical scores of some individuals in a problem situation. If a mechanism of motivation is actually defined by the concept, then the operation of the mechanism in an appropriate experimental situation should yield certain predictable results. Such predictions were made and if supported by the data would indicate that a mechanism of motivation was operative in the experimental situation which was used.

Definitions

Most tasks can be completed by a number of different methods (1). Many of these methods are of unequal difficulty (2) so that the individual is limited by his general ability as to the kinds of methods which he can use. An individual who is confronted with a task which can be completed by a number of methods of unequal difficulty may use the most difficult methods which are available to him in relation to his general ability to perform the task, or he may complete the task by methods which are far below his general ability.

Whether a person will use the most difficult methods which are available to him in completing a task, or whether he will use less difficult methods to obtain the same result is an expression of the *degree of effort* which he is exerting. If an individual uses the most difficult methods in relation to his ability in performing a task, he is exerting a *maximum degree of effort;* if an individual uses the less difficult methods to complete a task, although the more difficult methods are available to him, he is exerting a *minimum degree of effort.*

Antecedents of Problem

To test the relationship between general ability and the methods which are used to complete a task, a problem was used which could be completed by a number of different methods. It was found (2) that some of the methods which were used to complete the task depended on the general ability of the subject. Many exceptions, how-

[1] From the Department of Nervous and Mental Diseases. A report read at the American Psychological Association meetings, Sept., 1941.

ever, were present. Thus some subjects with a mental age level above 15 years used methods to complete the problem which characterized individuals below the mental age of 8 years. To explain these discrepancies it was postulated that the individuals who use the less difficult methods to complete the task are exerting a minimum degree of effort, whereas the individuals who use the more difficult methods, although the nature of the problem is such as not to require this to be done, are working at a maximum degree of effort.

Conditions Which Must Be Satisfied

Before the hypothesis that individuals differ as to the degree of effort which they exert can be accepted, it is first of all necessary to show that the more difficult solutions are available to those individuals who do not use them. In the second place, certain predictions (which were made before the experiment was initiated) concerning the individual's performance in the experimental situation must be satisfied. If all or at least a majority of the predictions are not satisfied by the data then the hypothesis of the degree of effort must be rejected.

Methods and Subjects

Apparatus.—The apparatus consists of a platform, 14 inches long and 4 inches wide, with a grooved runway extending the length of the board. The grooved runway holds an iron bearing, $\frac{3}{4}$ of an inch in diameter, with a hole, $\frac{3}{16}$ of an inch in diameter, drilled through it. A pendulum, which if released can be made to strike the bearing, is mounted at one end above the runway. A stick (the end of which can be inserted in the hole drilled through the bearing), a pair of pliers, and a magnet are mounted on a side board attached to the platform.

Open Runway Test.—On the first part of the test (hereafter designated as the open runway test) during each trail the bearing is placed in the end of the runway underneath the pendulum with the hole facing upward and visible to the subject. The subject is instructed to use any method, including the tools, to move the bearing down the runway. On the following trails the bearing is replaced by the experimenter in its starting position and the subject is told to move the bearing down the runway using different methods from those which he had already used. The test is continued until either 10 trials are completed or until the subject fails to complete a trial in 3 minutes. Preliminary tests had shown that if the subject does not complete the trial in 3 minutes he will not ordinarily do so if the time limit is extended. The time and the method used to move the bearing are recorded for each trial.

Blocked Runway Test.—When the test with the open runway is completed the bearing is placed in its starting position and a barrier is placed in the center of the runway. The subject is again instructed to move the bearing from its position of rest to the opposite end of the runway, but without removing the barrier, and by using any method which he wants to use, including the methods which he has used before. On the subsequent trials the subject is told to use a different method from that which he has already used on this part of the test. Five trials are given with the same time limit as in the open runway test. This procedure will be referred to as the blocked runway test.

Subjects.—Forty patients, 26 males and 14 females, with minor ailments such as a cold, recheck on an operation, etc., from the Northwestern University Medical Clinic, were used as subjects. The average age of the subjects was 25.8 years, with a range from 11 to 47 years. The subjects represent a fairly random sample of the population since the only selection which was used was that the patient have a facile use of the English language, that he be below the age of 60 years, and, because of the nature of the test, that his mental age rating be above 8 years.

Interpretation of Methods

On the open runway test it had been previously established (2) that there are three methods of completing the task which are related to the mental age levels of the subjects. They are, in the order of their difficulty, first, to apply some external force to the bearing (to push it with one's fingers, the pendulum, pliers, magnet, or stick) or to carry the bearing with one's fingers, second, to use the tools in their appropriate ways (to insert the stick in the hole drilled through the bearing, to clasp the bearing with the pliers, and to use the magnet), and third, to move the bearing by tilting the platform. The mental age levels dividing the subjects into the three groups were approximately 8 and 13 years. For reasons which will soon become evident only the methods involving the first two levels of difficulty, that of applying an external force to the bearing or carrying the bearing with one's fingers and of using the tools are considered in this study. The closed runway problem can be solved by carrying the ball over the barrier with one's fingers, the stick, the pliers, and the magnet.

The open runway problem allows the subject to perform the task either using a maximum degree of effort, since the subject may use all of the more difficult methods which are available to him, or using a minimum degree of effort during each trial, that of carrying the bearing with one's fingers or of applying an external force to the bearing by pushing it with the different tools, the different fingers, etc. The closed runway problem, however, can be solved only by the more difficult method of using the tools in their appropriate ways. *The open runway test shows whether the subject is working at a minimum or a maximum degree of effort. The closed runway test indicates whether the more difficult methods of completing the task are available to the subject.* If the individual fails to use one or more of the tools on the open runway test which he uses on the blocked runway test, he is working at a minimum degree of effort. Twenty-five of the 40 subjects fell into this classification. If the subject uses the same tools on the open and blocked runway tests he is working at a maximum degree of effort. Fifteen of the subjects satisfied this criterion. If the individual does not use the tools during either one of the tests, these methods of solution are not available to him. Such subjects were not used in the experiment.

Results

Predictions.—The predictions which were made concerned only the scores of the open runway test, since only this part of the test allowed the subjects to exert different degrees of effort. The in-

dividuals who exerted a maximum degree of effort were predicted to differ from the individuals who expended a minimum degree of effort in the following ways:

Prediction 1. They should take on the average a longer time to complete the test. Reason: The individual who is attempting to use the best, *i.e.*, more difficult, methods of completing the task would be more likely to deliberate on the problem, than would the individual who is satisfied to complete the task with inferior methods.

Prediction 2. They should take a longer time to finish the last half of the test. Reason: The individual who is striving to use the more difficult methods would be more likely to spend more time in seeking such methods as they become depleted upon successive trials, than would the individual who is satisfied to use the less difficult methods of completing the task.

Prediction 3. They should use the tools less frequently on latter part of the test. Reason: The individual who insists on using the more difficult methods would soon exhaust these methods and would be obliged to use the less difficult methods on the latter part of the test. This would not be true of the individual who is satisfied to use the less difficult methods throughout.

Prediction 4. They should discontinue the task more frequently. Reason: As the number of solutions become depleted the individual who insists on using the more difficult methods of completing a task would be more likely to discontinue the task when the more difficult methods are exhausted than would the individual who is satisfied to use the less difficult methods.

Table 1 summarizes the data on the first 3 predictions. For prediction 1 the average time to complete the trials was obtained. For prediction 2, in order to see if there is a significant increase in the time to complete the latter half of the test, the average time to

TABLE 1

AVERAGE TIME IN SECONDS TO COMPLETE THE OPEN RUNWAY TEST (PREDICTION 1), AVERAGE TIME SPENT ON THE FIRST HALF OF THE TEST SUBTRACTED FROM THE AVERAGE TIME SPENT ON THE SECOND HALF (PREDICTION 2), AND THE NUMBER OF TIMES THE TOOLS WERE USED ON THE FIRST HALF AND THE SECOND HALF OF THE TEST (PREDICTION 3) OF THE GROUPS EXPENDING A MAXIMUM AND A MINIMUM DEGREE OF EFFORT

	Prediction 1. Ave. Time to Complete Test	Prediction 2. Difference (1st.–2nd half)	Prediction 3. No. of Tools Used		
			1st half	2nd half	χ^2
Maximum degree of effort	23.4	23.4	31	15	5.6
Minimum degree of effort	11.7	7.8	14	11	0.3
t test	3.25	2.44			

complete the trials on the first part of the test was subtracted from the average time taken to complete the second half. The number of trials were split into two equal parts for each subject even if all of the 10 trials were not completed. The *t* test shows that the differences between the two groups on prediction 1 and 2 are significant. The validity of prediction 3 can be determined by testing whether the tools are used with the same frequency by each group on the first half and the second half of the test. The 50 : 50 hypothesis, using the chi-square test, applied to these data shows that the individuals who expend a maximum degree of effort use the tools less frequently on the second half of the test. This is not true of the individuals who expend a minimum degree of effort. The first 3 predictions are, therefore, substantiated by the data. The fourth prediction, however, is not since 5 of the 15 subjects in the maximum degree of effort group and 4 of the 25 subjects in the minimum degree of effort classification failed to complete the test. The probability that this difference is significant is only .20.

General Ability.—If the degree of effort deals with motivation, there should be no difference between the general abilities of the individuals displaying a maximum and a minimum degree of effort. Intelligence tests [2] were administered to 14 subjects in the maximum degree of effort group, and to 19 subjects in the minimum degree of effort classification. The average mental age of the former group was 13.4 years and of the latter 13.1 years with a *t* value of 0.66. There were also no chronological age differences or sex differences between the two groups of subjects.

Discrete Versus Continuous Distribution.—The method which was used to measure the degree of effort was such that the subjects could be divided most readily into two categories. One would expect, however, judging from the distribution of traits in general, that this characteristic would follow a continuous distribution, rather than being divisible into two categories.

In order to test this possibility the subjects composing the minimum degree of effort group were divided into three sub-groups on the basis of the number of tools which they failed to use on the open runway test but used on the blocked runway test. This procedure was used since the individual who fails to use all three tools might be expected to be working at a lower degree of effort than the individual who fails to use only two tools; and the latter individual

[2] The Stanford-Binet, Form L, or the Wechsler-Bellevue intelligence tests were administered to the subjects before they were given the test for the degree of effort. Since the number of cases in each group taking the two different tests was almost proportional, the error, for comparative purposes, involved in combining the scores of two such divergent tests should be negligible.

might be expected to be expending less effort than the individual who fails to use one tool. An individual was considered to fail to use a tool on the open runway test only if this tool was used by him when the runway was blocked. Finally, only the first five trials on the open runway test were used to make this classification, *i.e.*, the tools used after the fifth trial were disregarded.[3] The classification of the subjects into the four grops, those failing to use 3, 2, and 1 tools on the open runway test, and those not failing to use the tools, gave the following number of subjects in each group: 8, 14, 8, and 10.

The data of prediction 1, 2, and 3 (Table 1) were tabulated into these four divisions and are shown in Fig. 1. All of the three meas-

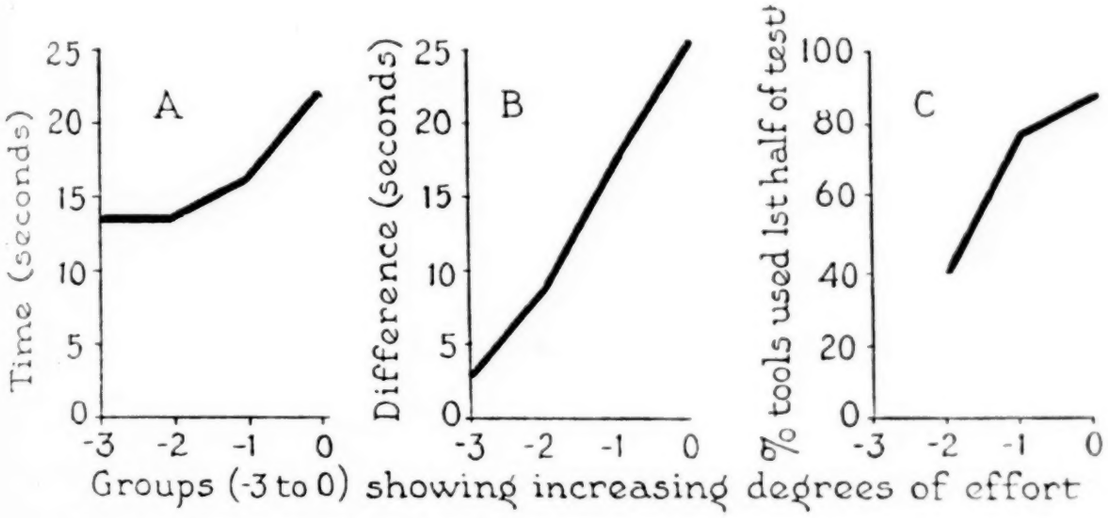

FIG. 1. The −3, −2, −1, and 0 on the abscissa represent the groups expending an increasing degree of effort (the number of tools used on the blocked runway test subtracted from the number of tools used on the open runway test). Figure *A* (from prediction 1) gives the average time taken for each trial to complete the test. Figure *B* (from prediction 2) gives the average time taken to complete the first half of the test subtracted from the time taken to complete the second half of the test. Figure *C* (from prediction 3) shows the number of trials on the first half of the test during which the subjects used the tools expressed as a percentage of the total number of trials of the whole open runway test during which the tools were used. For group −3 of this figure the score is indeterminate since the subjects did not use any of the tools during the whole test.

ures which originally differentiated the individuals working at a maximum or a minimum degree of effort are so distributed between these four groups as to indicate that the degree of effort follows a continuous distribution. The scores of prediction 4, as might be expected since the prediction was not substantiated by the data, did not follow this distribution.

[3] From prediction 3 (Table 1) it appears that the latter half of the open runway test adds little to the classification of the individuals as to the degree of effort which is expended, and might even give some individuals a higher score than the degree of effort they are exerting warrants, since they would have more opportunity to use the tools in a haphazard fashion without actually attempting to use the more difficult methods of completing the task.

Discussion

If an individual is given a task which can be completed by a number of methods of unequal difficulty he can complete the task by using either the less or the more difficult methods. The thesis has been advanced that the difficulty of methods which an individual uses is an expression of the degree of effort which he is exerting.

The test situation was such that certain predictions concerning the individual's scores could be made, and should be verified by the experimental results if the individuals actually differ in the degree of effort which they exert to complete the task. Three of the four predictions which were made were verified by the results. A characteristic defined as the degree of effort was, therefore, operative in the context of the experimental situation which was used.[4]

Further evidence that a characteristic in the field of motivation is being measured is offered in the fact that there is no relationship between the degree of effort and general ability, and that the degree of effort appears to be distributed on a psychological continuum. The present conception of motivation and traits in general demands that these two criteria be satisfied.

Summary

1. A behavior characteristic differentiating individuals on the basis of the difficulty of methods which they use to complete a task has been defined as the degree of effort. If a task can be completed by a number of methods of unequal difficulty the individuals who use the more difficult methods are exerting a maximum degree of effort; the individuals who use the less difficult methods are exerting a minimum degree of effort.

2. On the test situation which was used to differentiate the two groups it was predicted that the individuals exerting a maximum degree of effort would differ from the individuals expending a minimum degree of effort in the following ways: (1) They should take a longer time to complete the test. (2) They should take a progressively longer time to complete the test as the more difficult methods become depleted. (3) They should use the less difficult methods only after the more difficult methods are used up. (4) They should discontinue the task when the more difficult solutions are exhausted. The first three predictions were verified by the results; the fourth prediction was not.

3. There is no relationship between the degree of effort and the mental age, chronological age, or sex.

[4] Two articles to follow deal with the broader implications of the operation of this characteristic.

4. The degree of effort characteristic appears to be distributed on a psychological continuum.

5. It is concluded that a motivational characteristic defined as the degree of effort was operative in the experimental situation which was used.

(Manuscript received July 6, 1941)

References

1. Seashore, R. H., Work methods: An often neglected factor underlying individual differences, *Psychol. Rev.*, 1939, **46**, 123–141.

2. Yacorzynski, G. K., An evaluation of the postulates underlying the Babcock deterioration test, *Psychol. Rev.*, 1941, **48**, 261–267.

CHANGES IN MUSCULAR TENSION DURING LEARNING *

BY C. W. TELFORD AND WALTER J. SWENSON

University of North Dakota and Bismarck Junior College

Introduction

There seems to be some conflict between the beliefs of teachers and coaches concerned with the direction of learners in the field of motor skills, and the findings of experimentalists as to the role of muscular tension in the development and performance of these skills. Instructors, for the most part, have advised relaxation on the part of the performer. The golf professional, the tennis coach, the piano teacher, and the penmanship instructor have all insisted that relaxation is conducive to rapid learning and good performance. Observations of the novice seem to disclose generalized and exaggerated muscular tensions. The child starting to learn to write grasps the pencil with all his strength; he holds himself rigid in his desk; he screws up his mouth, sticks out his tongue, winks frequently, breathes rapidly, hooks his feet around the legs of the chair, wrinkles his brow, maintains a strained expression on his face, and in all respects seems to present a picture of overtensions and exaggerated muscular strains.

As learning proceeds and the child becomes more proficient, these tensions apparently disappear. He sits easily in his chair, his face loses its strained expression, he winks less rapidly, his legs and feet hang freely and altogether he presents a picture of greater relaxation. From observations of this sort, we have been led to believe that as learning progressed muscular tension decreased, that the observed tensions were detrimental to the skilled performance and that the more completely they could be eliminated the more efficient would be the performance.

In contrast to this, we have a rather long series of experimental studies indicating that voluntarily induced muscular tensions facilitate both motor and mental as well as sensory processes. Breese (3) as early as 1899 found that increased muscular tensions induced by requiring the subjects to press their arms tightly against their sides increased visual acuity. Jacobson (22) likewise in 1912 found that increased muscular tension increased auditory acuity. These and

* The senior author was responsible for the general direction of the present study. He also reworked the data from the original study and prepared the manuscript for publication. Swenson gathered the data and made the necessary analyses. The more complete original study is available in the form of an unpublished thesis in the library of the University of North Dakota.

other similar studies indicate that induced muscular tensions produce lowered sensory thresholds.

Bills (**1**) in 1927 required his subjects to maintain a given tension by squeezing a dynamometer in the unused hand while they memorized four different types of ideational material. With all four types of material he found that the increased muscular tension facilitated learning. Further experiments by the same worker in 1937 gave conflicting results (**2**). Zartman and Cason (**30**) found a slight advantage for increased tension in solving arithemtical problems. Stroud (**28, 29**) found a positive correlation between degree of muscular tension and the difficulty of the task in maze running. With easy mazes he found that tension tended to increase with each successive trial. He found a positive correlation between speed and tension. He concludes that there is an increase in muscular tension during effortful activity and that these tensions are valuable in activities of the kind involved in maze learning. Johnson (**24**) likewise found efficiency of performance in tapping to be directly related to tension.

G. L. Freeman (**7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18**) has contributed a long series of valuable studies on the general question of the influence of muscular tensions of various kinds on the performance of a variety of mental and motor tasks. In the studies most closely related to the present one he has presented evidence to show that there are degrees of muscular tension which are optimal for the performance of particular tasks. Varying the degree of tension in either direction from this optimum results in decreased output. Courts (**4**) and Stauffacher (**27**) have also presented evidence in support of this same conception.

It should be mentioned that a few studies have failed to show a beneficial effect from increased muscular tension. Russell (**26**) found accuracy in ball tossing and degree of tension to be inversely related. Ghiselli (**19**) found that pressure on the response-key during multiple-choice learning was maintained at a fairly constant level. During over-learning there was a marked decrease in tension.

In summary, it seems that most of the experimental studies show a positive correlation between quality of performance and degree of muscular tension. This is in opposition to the prevalent opinion among instructors that muscular tensions interfere with these performances.

Problem

In the present problem we have determined the changes in muscular tension throughout the entire learning process as displayed in a typical case of sensorimotor learning. Previous studies have been

concerned largely with the relationship of degree of tension to quality of performance at a given level of performance, or they have sampled performance at various stages in the learning process. We have followed the level of muscular tension from the beginning of learning to a level of very high proficiency and noted changes in tension as the learning progressed. Most of the earlier studies were concerned with the effects of artificially or voluntarily induced tensions, whereas, the present study is concerned with changes in tension which occur when the subjects are left perfectly free to vary tension as they choose. As a matter of fact, all the subjects used were kept in entire ignorance of the purpose of the experiment.

Mirror drawing was selected as the learning problem and the tension measured was that exerted by the writing hand on a pneumatic stylus during the drawing.

Apparatus and Procedure

After considerable preliminary experimentation, a tracing pattern consisting of a path approximately one-fourth inch wide was constructed. The pattern was of irregular shape, with a series of straight and curved sections which changed directions several times and terminated near the starting point. This was found to constitute a fairly difficult and satisfactory problem for the purposes of the experiment. The pattern was cut in metal and insulated at the base with bakelite. The inside edges of the path were notched to prevent the pattern being traced by continued contact with either of the sides. The metal plate was connected in series with a dry cell, stylus, and recorder so that any contact with the edge of the pattern (constituting an error) would close the electric circuit and record on a kymograph.

The stylus used in tracing the pattern was of the pneumatic type, seven inches long and one and one-half inches in circumference. The stylus was made of a rubber tube closed at both ends by rubber stoppers which were cemented into place. Throughout the length of the rubber tube ran a quarter-inch brass tube, perforated in several places to admit air from the pneumatic stylus when it was compressed by the subject. A hard rubber tube of small bore connected the upper end of the stylus with the tambour which, in turn, recorded changes in tension on a kymograph. At the lower end, the brass tube was plugged by an L shaped tracing end. The stylus was marked in two places, one on either side, in the position commonly assumed by the thumb and index finger in writing or tracing, and the subject was always required to maintain these positions. This was done in order to minimize differences in tension due to different techniques used in holding the stylus.

The recording apparatus, shielded from the view of the subject, was arranged at one end of a well-lighted table of suitable height. The experimenter and subject were seated at opposite ends of the table. Conditions of light, temperature, and time of day were kept constant throughout the experiment.

Subjects.—Fairly complete records were obtained from twenty-five subjects. These consisted of two sub-groups. One of these was made up of graduate students attending the University of North Dakota and members of a city school faculty. A small group of children from the grades and high school made up the second sub-group. All subjects were average or above in intelligence and coöperated well throughout the experiment.

Procedure.—The problem was presented to the subjects as a simple learning task. The apparatus used in the tracing was shown, the method of holding the stylus was demonstrated, and the importance of uniformity of procedure, particularly the method of holding the stylus, was emphasized.

During the tracing, any contact with the side of the path activated the recorder and produced an audible click, thus giving the subject knowledge of his errors. Simultaneous records of tension were obtained on the kymograph. The time required for each tracing was measured by means of a stop-watch.

One hundred tracings were obtained from each subject. Fifteen trials were taken daily for six days and the remaining ten trials were completed on the seventh day. With nine of the subjects, following the one hundredth trial and at the same sitting, an additional hour was spent in taking tracings in order to discover the possible effects of fatigue and increased motivation on errors and tension.

In the beginning, tension readings were made for each of the one hundred trials. However, the amount of time required to take the readings on this number of trials did not seem commensurate with the returns. Consequently, subsequent tension readings were taken on each fifth trial instead of every one. The tension records were measured in millimeters above the base line, separate measurements being made at quarter-inch intervals. Notes were always made of any unusual emotional, mental or physical states of the subjects which were present prior to, or which developed during the course of the sitting.

Results

In examining the tension patterns for each trial there appear some interesting regularities. In the earlier runs tension is low at the beginning of the trial and increases irregularly but fairly constantly until it is highest at the end of each trial. As learning progresses the initial tension increases and remains on a more constant level to the end of the trial until in the last trials there is but little difference between the tension at the beginning and the end of the run. These differences are shown in the sample curves reproduced in Fig. 1.

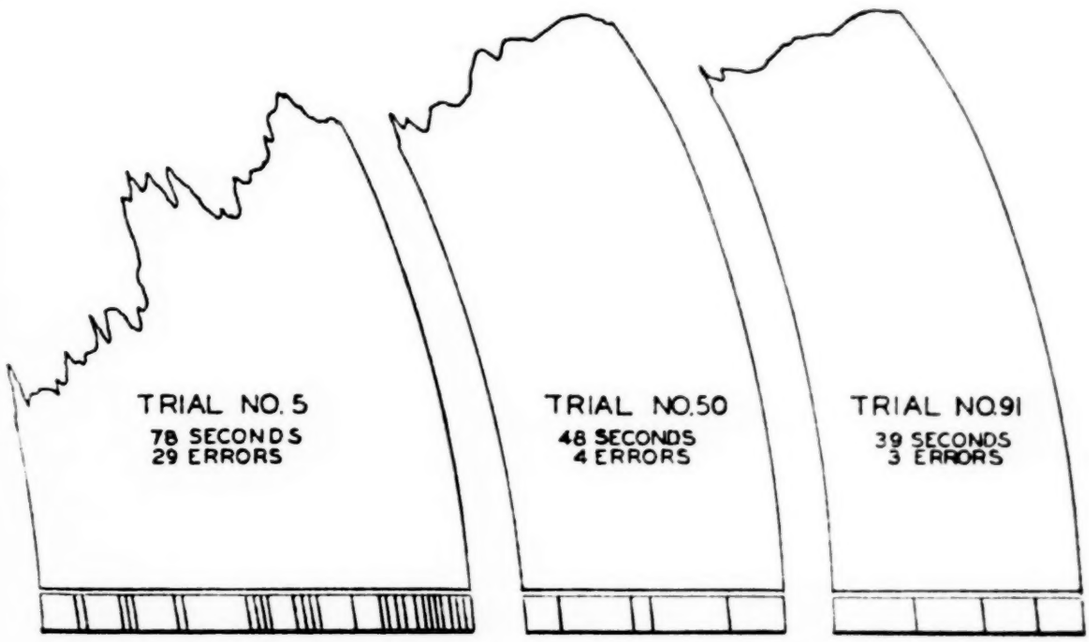

Fig. 1. Sample tension records from same subject.

An examination of these individual curves also shows that as learning progresses the tension exerted becomes much more constant.

Figure 2 shows the changes in average tension, errors, and time for each fifth trial for all of the twenty-five subjects.

The curves for time and errors are, of course, typical for learning of this type. The tension curve shows an initial decline which con-

tinues for the first 15–20 trials, then it rises irregularly to the end of the learning. The initial decrease in tension coincides pretty well with the initial rapid decrease in time and errors.

An examination of the twenty-five individual curves shows that, with one exception, they all follow the same general trend to the

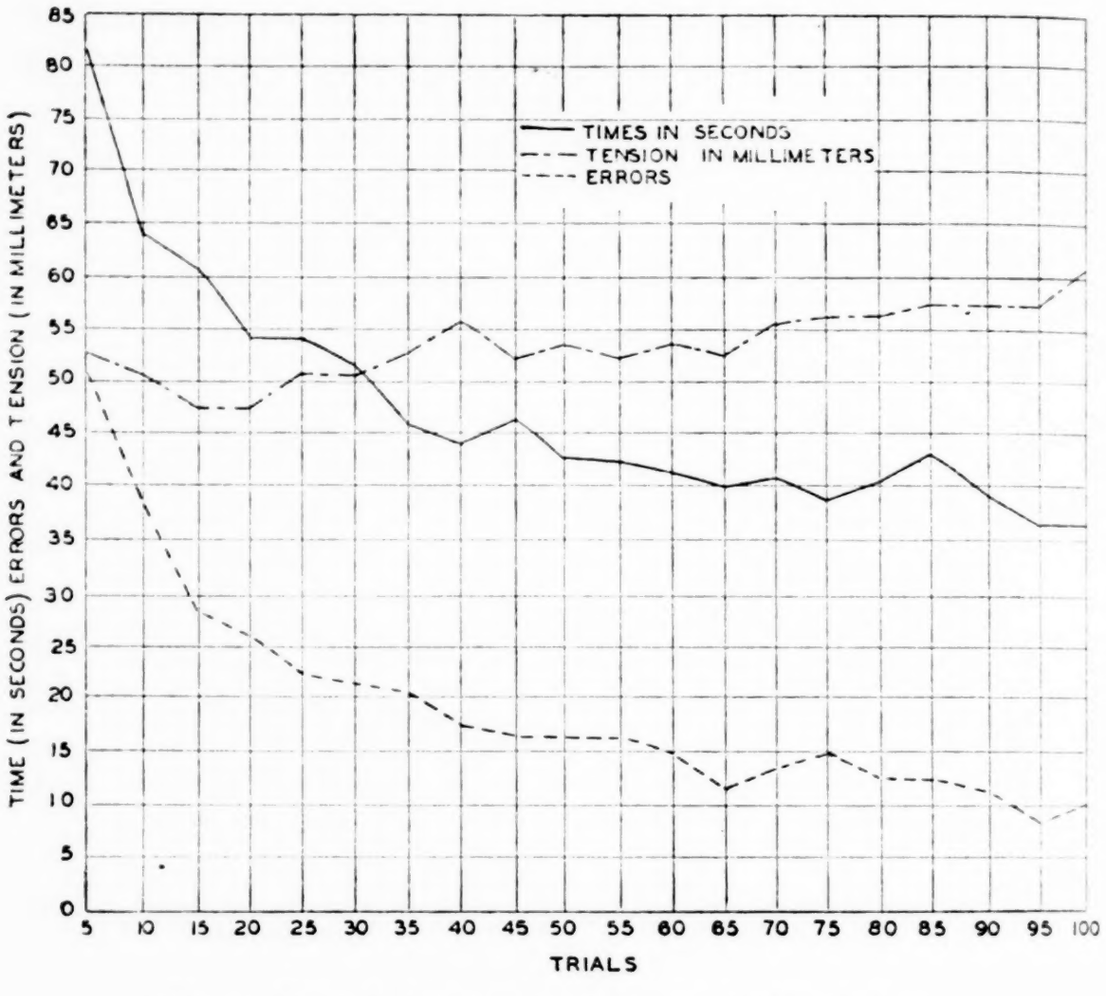

Fig. 2. Group time error and tension curves in mirror drawing.

extent of showing an initial decline in tension followed by a period of irregular fluctuations and a subsequent rise to the end of the series. Although the particular points of change in the general direction of the tension curves varies considerably with the different subjects, they all follow the same general pattern.

Because of the nature of the curves, correlational techniques are not particularly applicable to these data, but we have computed the correlations between tension and errors as well as between tension and time for the last half of the learning (the last fifty trials). All of the individual curves were beyond the initial decline by the fiftieth trial.

Twenty-four of the twenty-five correlations between tension and errors are negative and only one is positive. The average of the individual correlations (Fisher method) is −.48. The average of the individual correlations between time and tension for the last fifty trials is −.63. Computing the correlations on the basis of the group averages as shown in Fig. 2 we get a correlation of −.63 between errors and tension and −.61 between time and tension for the entire one hundred trials.

Relation of Points of Particular Difficulty to Tension.—In constructing the tracing pattern an irregularly shaped figure was constructed with no particular effort being made to keep the difficulty of different parts of the path equal. From inspection of the finished

FIG. 3. Sample tension record with arrows indicating points of increased difficulty in tracing pattern.

pattern, as well as the records of errors made and comments of the subjects during the learning, it was evident that three points in the pattern offered special difficulties. For several of the subjects, the points in the tension curves corresponding to these places in the pattern were marked on the kymographic records. A typical record showing these points is shown in Fig. 3.

The arrows indicate the regions of the tension curve corresponding to the points of exceptional difficulty in the tracing. It will be noted that these places are marked by decided increases in tension. Apparently increased difficulty within a pattern was met by increased tension on the part of the subject. This is in keeping with Davis' (**6**) finding that there occurred an increase in muscular tension during the solution of a problem corresponding in amount to the difficulty of the problem. McTeer (**25**) likewise found that punishment for

errors tends to produce greater tension in the inactive hand during the course of tracing activity.

Influence of Individual Errors on Tension.—In studying the individual tension curves together with notes made at the time of the tracing some relationships worth mentioning were noted. It was noticeable that the increase in tension was especially marked in individual runs where points of difficulty were approached after few or no errors had been made previously. After traversing most of the pathway without errors the tension curve would typically show an increase and remain high to the end of the trial. In contrast to this, the records which showed an abnormal number of errors in the early parts of a trial characteristically showed decreases in tension as the trial progressed. This was also usually accompanied by an increasing tendency to make additional errors.

In the regular runs, where about the normal number of errors is being made, there is a slight increase in tension subsequent to the making of each error.

Effects of Emotion on Tension.—On two occasions during the learning the subjects had been angered just before the day's runs began and still showed some emotional disturbance during the session. In both cases the cause of the irritation was outside the laboratory but the subjects were still rather disturbed at the time of their learning. Their curves show marked changes in the direction of increased irregularity and an absence of changes in degree of tension with the varying difficulties of the different parts of the pathway. The normal increases in tension following the occurrence of errors were also strangely absent from the curves. If anything, there was a slight tendency for the tension to decrease rather than increase subsequent to the errors. There was also a decided tendency for tension to decrease from the first to the fifteenth trial of the sitting. This may have been due to a lessening of the emotional state.

One other subject, who showed considerable elation at the time of one of the sittings, likewise showed an absence of the usual relationship between errors and tension as well as between points of difficulty in the pattern and tension. The tension was also abnormally irregular. The average tension was maintained at a very high level throughout the sitting.

Influence of Fatigue and Excessive Urging on Tension.—Nine of the subjects serving in the main part of the experiment took part in a small supplementary experiment which followed the completion of the original learning. These subjects spent one hour in making additional runs at the completion of the original one hundred trials. In these 'fatigue' runs the subjects were urged to do their best. In

all of these runs the tension was considerably higher than in the earlier regular runs and tended to increase from the beginning to the end of each trial. Tension also increased in successive runs from the beginning to the end of the hour. The influence of errors on the tension curve is also less evident. This suggests that the subjects were concentrating more on speed and were paying less attention to errors. The correlations between errors and tension, however, are about as they were for the regular trials. All of the correlations were negative except one, the average of the individual correlations being $-.51$.

Excessive urging on the part of the experimenter produced some noticeable changes in the tension curves. The curves were much more regular, the average tension was abnormally high and showed a constant rise from the beginning to the end of the trial. The effect of errors on tension is also exaggerated in these runs.

Age and Tension.—The tension curves of the younger subjects tended to be somewhat higher throughout the learning than those of the adults. The shape of their curves and the general trend of their results were in all other ways comparable to those of the adults.

Interpretations and Conclusions

In general our results suggest an answer to the apparent discrepancy between the preachments of teachers who are concerned with the teaching of motor skills and the findings of the experimentalists as to the role of muscular tension in the acquisition and performance of these skills. The naïve adult's approach to the mirror-drawing problem is in many ways similar to that of the child in learning to write. When confronted with the new and puzzling situation both react with widespread and diffuse motor tensions. Our experimental set-up was not such as to demonstrate objectively the existence of tensions other than those in the writing hand. But observations of the postures, facial expressions, rate of winking and respiration as well as the evidence of other experimenters indicate that such widespread tensions do exist. (See Freeman **7**, **10**, **12**, **14**, and Davis **5**.) As the learning proceeds and the learner makes the preliminary adjustments to the general situation the generalized tensions tend to decrease. This occurs in the mirror-drawing situation when the subject, besides becoming accustomed to the general learning situation, begins to 'get the hang' of how to reverse the apparent directions as he watches his hand in the mirror and is reasonably successful in getting around the pathway. As this adjustment progresses, tension decreases along with time and errors. The initial decrease in the tension curve found in our data is a part of the reduction of the excess generalized tension from which the individual suffers.

The initial reduction in tension is followed by a period of increased localization of tension in the muscle groups specifically involved in the task. This increased localization shows itself in a rise in the tension exerted by the hand holding the stylus. As the process becomes more habituated the general neuro-muscular activity and resulting tensions are reduced and the pattern becomes more specifically localized in the organs involved in the execution of the acts being learned or performed. Freeman (**7**) found this to be true of pursuit learning. He recorded muscular tensions from four muscle groups simultaneously and found that in the beginning the performance was poor and the muscular tensions were high in all of the four groups under survey. With the acquisition of skill, the neuro-muscular activity tended to be more and more localized in the muscle groups specifically involved in the task. In the same study he presented evidence which showed that in the learning of successive lists of nonsense syllables there was a reduction in quadriceps tension as a consequence of practice in learning nonsense syllables. This is probably comparable to the reduction in general tension found in the early stages of learning in our experiments.

There is, of course, a certain economy in this. Tensions in the back, legs, face and other remote regions of the body do not contribute to the tracing of the pattern. In fact, they may interfere with the process. There is probably considerable tension in the antagonists of the muscles specifically involved in the tracing. This represents just so much additional force to be overcome in making the desired responses and tends to interfere with the proper muscular coordinations. As these diffuse tensions are released a great deal of waste energy is conserved. The interferences from conflicting tensions gradually drop out and this likewise facilitates the proper motions. Along with, or following, the release of these superfluous tensions, there occurs a concentration of tension in the particular muscle groups which are used in tracing. In other words, what we have is a development of a set of specialized tensions superimposed upon a condition of more widespread relaxation.

Thus, it is probable that the teacher of motor skills has had in mind the widespread, diffuse, overtenseness which characterizes the novice, when he advised relaxation as conducive to good performance. What he really desires, however, is not complete but a differential relaxation. He wants the tensions in the muscles involved in the performance of the specific response maintained or even increased but he also wants the generalized interfering tensions released. Of course, what we need for the most efficient performance is a continuously changing pattern of muscular tensions as the demands of the situation vary and an absence of interfering tensions with only

sufficient generalized postural support to maintain the activities in progress. Anything more than this will result in a loss of efficiency.

It is obvious from this, as Freeman (**11, 12**) and Stauffacher (**27**) have shown, that for each activity and each individual there is a degree of generalized and localized tension which is conducive to maximum efficiency. Tensions which are either greater or smaller than this will interfere with maximum performance. Stauffacher found that in his learning problems, poor learners benefited from induced muscular tensions while the best learners did not. This suggests that the best learners had found and maintained an optimal degree of tension for good performance whereas the poorer students maintained tensions which were in general below the maximum for efficiency. Jacobson (**21** and **23**) has shown that when the subject is too relaxed, mental work, if it occurs at all, is on a very low level.

Summary

In the present study we have shown the following:

1. For fairly difficult mirror-drawing, muscular tension exerted by the hand holding the stylus in tracing at first decreases and then increases irregularly as the learning progresses.

2. Taking the last fifty trials as the point beyond which the change in direction of the curve has taken place in practically all of the individual curves we find average correlations of $-.48$ between errors and tension, and $-.63$ between time and tension based on the individual data. When the averages of all twenty-five subjects for the entire one hundred trials are taken as the basis for correlation the values are $-.63$ for errors and tension and $-.61$ for time and tension.

3. Points of increased difficulty in the tracing pattern were associated with increased tension on the part of the subjects in attacking these points.

4. A good performance in the first parts of a single run was usually followed by high tension throughout the rest of the trial, whereas records characterized by an abnormally large number of errors in the early parts of the run were characterized by decreasing tension and an increasing number of errors during the remainder of the trial.

5. Normally the occurrence of an error resulted in a slight increase in tension.

6. Emotion seemed to increase the irregularity of tension and eliminate the normal increases in tension subsequent to errors and at the points of increased difficulty in the tracing pattern.

7. Fatigue and excessive motivation increase average tension. They do not change the relationship between tension and errors.

8. Our results are interpreted as indicating that in the learning process there occurs a reduction in the initially high generalized diffuse tension with a subsequent increased localization of tension in

the muscle groups specifically involved in the performance of the task at hand.

(Manuscript received July 18, 1941)

References

1. Bills, A. G., Influence of muscular tension on mental work, *Amer. J. Psychol.*, 1927, 38, 227–240.
2. Bills, A. G., and Stauffacher, J. C., The influence of voluntarily induced tension on rational problem solving, *J. Psychol.*, 1937, **4**, 261–271.
3. Breese, B. D., On inhibition, *Psychol. Monog.*, 1899, **3**, No. 31.
4. Courts, F. A., Relation between experimentally induced muscular tension and memorization, *J. exp. Psychol.*, 1939, **25**, 235–256.
5. Davis, R. C., Induced shifts of muscular tensions during mental work, *Psychol. Bull.*, 1938, **35**, 667–668.
6. Davis, R. C., Relation of muscle action potentials to difficulty and frustration, *J. exp. Psychol.*, 1938, **23**, 141–158.
7. Freeman, G. L., Spread of neuro-muscular activity during mental work, *J. gen. Psychol.*, 1931, **5**, 479–494.
8. Freeman, G. L., Muscular tension and inhibitory effect of muscular tension upon performance, *Amer. J. Psychol.*, 1932, **44**, 789–792.
9. Freeman, G. L., The facilitative and inhibitory effect of muscular tension upon performance, *Amer. J. Psychol.*, 1933, **45**, 17–52.
10. Freeman, G. L., The optimal locus of anticipatory tensions in muscular work, *J. exp. Psychol.*, 1937, **21**, 554–566.
11. Freeman, G. L., The optimal muscular tension for various performances, *Amer. J. Psychol.*, 1938, **51**, 146–150.
12. Freeman, G. L., Our muscles and our minds, *Sci. Mo.*, 1938, **47**, 161–167.
13. Freeman, G. L., and Simpson, R. M., The effect of experimentally induced muscular tension upon palmar skin resistance, *J. gen. Psychol.*, 1938, **19**, 319–326.
14. Freeman, G. L., The postural substrate, *Psychol. Rev.*, 1938, **45**, 324–334.
15. Freeman, G. L., The effect of inhibited micturition upon interrupted and completed acts of unrelated origin, *J. gen. Psychol.*, 1938, **19**, 277–283.
16. Freeman, G. L., Postural accompaniments of voluntary inhibition of micturition, *J. exp. Psychol.*, 1938, **23**, 45–61.
17. Freeman, G. L., Changes in tension-pattern and total energy expenditure during adaptation to 'distracting' stimuli, *Amer. J. Psychol.*, 1939, **52**, 354–360.
18. Freeman, G. L., and Kendall, W. E., The effect upon reaction-time of muscular tension induced at various preparatory intervals, *J. exp. Psychol.*, 1940, **27**, 136–148.
19. Ghiselli, E., Changes in neuro-muscular tension accompanying the performance of a learning problem involving constant choice-time, *J. exp. Psychol.*, 1936, **19**, 91–99.
20. Henley, E. H., Factors related to muscular tension, *Arch. Psychol.*, 1935, **27**, No. 183, pp. 44.
21. Jacobson, E., Electrophysiology of mental activities, *Amer. J. Psychol.*, 1932, **44**, 677–694.
22. Jacobson, E., Further experiments on inhibition of sensation, *Amer. J. Psychol.*, 1912, 23, 345–369.
23. Jacobson, E., *Progressive Relaxation* (Rev. Ed.), Chicago: Univ. Chicago Press, 1938.
24. Johnson, B., Changes in muscular tension in coördinated hand movements, *J. exp. Psychol.*, 1928, **11**, 329–341.
25. McTeer, W., Changes in grip tension following electric shock in mirror tracing, *J. exp. Psychol.*, 1933, **16**, 735–742.
26. Russell, J. T., Relative efficiency of relaxation and tension in performing an act of skill, *J. gen. Psychol.*, 1932, **6**, 330–343.
27. Stauffacher, J. E., The effect of induced muscular tension upon various phases of the learning process, *J. exp. Psychol.*, 1937, **21**, 26–46.
28. Stroud, J. B., The role of muscular tension in stylus maze learning, *J. exp. Psychol.*, 1931, **14**, 606–631.
29. Stroud, J. B., Apparatus for measuring muscular tension, *J. exp. Psychol.*, 1931, **14**, 184–185.
30. Zartman, E. W., and Cason, H., The influence of an increase in muscular tension on mental efficiency, *J. exp. Psychol.*, 1934, **17**, 671–679.

AN EXPERIMENTAL STUDY OF GUTHRIE'S THEORY OF REINFORCEMENT

BY JOHN P. SEWARD

Connecticut College

The concept of reinforcement in learning theory stems from the fact that only under certain conditions does a neutral stimulus become capable of arousing a concurrent response. Reinforcement may be defined as the process by which a response to a situation is more readily aroused when some feature of the situation recurs. In spite of its fundamental importance the nature of this process remains unknown, although the most comprehensive discussion of the problem to date (**7**, 75 ff.) shows it to be the subject of active theoretical controversy.

The most arresting theory of the mechanism of reinforcement is that of Guthrie; arresting because, being couched in stimulus-response terms, it appears susceptible of experimental verification or refutation. It might be called the *finality theory of reinforcement.* Briefly, Guthrie holds that of the various responses made to a disturbing situation the final response is the one which is most likely to be repeated the next time the situation is met. In the author's own words:

> The next time that the disturbers are present they will tend to call out, by virtue of their last association, the act that removed them. Other acts associated with them have been dissociated or unconditioned each by the next act. But after successful removal of the disturber, *it is no longer there to be associated with a new act.* The drive remains faithful to the act that removed it because that was its last association (**4**, 98).

Indeed, Guthrie maintains that the sole effect of reward on learning is to prevent further responses to the annoying stimuli. Of a cat in a problem box he writes: "*The learning took place on the first performance before escape or reward. All that escape or reward does is to protect the learning from being unlearned* (**3**, 484)." [1]

The present experiment was designed to subject this theory to experimental test; *i.e.*, to determine whether reward has any effect on learning apart from finality, and, conversely, whether finality

[1] Holt (**8**) also emphasized the importance of finality in the learning of avoidance responses. He did not rely on it as the sole principle of reinforcement, however, but held that in many trial-and-error situations the successful response was strengthened by intense afferent stimulation due to release of the inhibited consummatory reaction.

has any effect on learning apart from reward.[2] Essentially it provided a comparison between the strengths of the final response to a situation under two conditions; when rewarded by food and when not so rewarded.

Method

Two experiments were performed, differing only in details; details, however, which precluded lumping all of the data. In general, the procedure was as follows: Two groups of rats were trained in a bar-pressing situation. In one, the reward group, the rat received a pellet of food upon pressing the bar and was then removed from the box. In the other, the removal group, it was simply lifted out of the box as soon as it had pressed the bar. In all other respects except the pellet the two groups were treated alike.

Two precautions in particular were taken, one before, one after each trial. Just before being put in the box each rat was fed one pellet. This was done in order to minimize the difference in internal stimulation between the two groups introduced by ingestion of the reward. The preliminary pellet presumably stopped hunger contractions of the stomach, according to the indirect evidence of Elliott and Treat (1). Otherwise it might be argued that in stopping these contractions the reward terminated one source of stimulation which mere removal did not, thus preventing a crucial test of the finality theory.[3]

The second precaution was to delay the animals' return to their living cages. An immediate return might well have constituted a reward for both groups. In Experiment 1 they were placed in a detention cage for at least 15 minutes after each trial before being put back in their living cages and fed. It seemed possible that even the detention cage, being larger, darker, and more sociable than the experimental box, might serve as a reward. In Experiment 2, therefore, each rat was transferred from the experimental box to a very similar one, slightly smaller and bare of gadgets, where it was kept in solitary confinement for at least 15 minutes.[4] From here it went to the detention cage for another 15 minutes and finally to the living cage, where it received its daily ration.

The animals were allowed to eat enough to maintain their weight; food was removed from $18\frac{1}{2}$ to 22 hours before the next day's trial. The order of rats was rotated from day to day to equalize the subsequent detention periods.

The experimental box was of unpainted wood, $11 \times 10\frac{1}{2} \times 9\frac{1}{4}$ inches, with a detachable cover of hardware cloth through which the rat could be observed and by means of which it was lowered in and lifted out by hand. From one side projected a flat brass bar, bent twice like a right-angled Z. Through a hole beside it the pellets fell into a small pan below. A downward pressure on the bar was recorded on a wax-paper kymograph, which involved an audible click from a signal magnet. Only pressures by the forefoot or head were counted; pressures by the hind foot or tail were not considered responses to the lever as a stimulus.

In Experiment 1 eight rats were used, two males and two females in each group. They were given one trial a day for approximately three weeks of training (19 and 23 days for males and females, respectively), followed by one 15-minute extinction trial a day for three days.[5]

---

[2] In Mowrer's (10) opinion this is an academic issue, since, operationally, it makes no difference whether the finality or the reward is the effective factor. His statement implies that the two factors are experimentally inseparable; it would be true only if it could be shown that the theoretical propositions involving each led to the same deductions regarding behavior. Neither of these suppositions seems justified at present.

[3] Guthrie (3) was recently led by McCulloch's (9) criticism to place less emphasis on the inner drive stimulus and more on stimuli from the animal's own restless movements. The above argument, however, still carries weight.

[4] From the standpoint of the theory under consideration it should be noted that in Experiment 2 the effective stimuli—*i.e.*, those to which bar pressing was the final response—were reduced to the bar, food pan, hole, and other less obvious respects in which the experimental and detention boxes differed.

[5] In the training trials if a rat failed to press the bar in 15 minutes it was removed and the next day the bar was smeared with food. This happened rarely, with the exception of one sluggish animal in the removal group, and had no apparent bearing on the results.

In Experiment 2 there were eight rats in each group, four males and four females. They were given one training trial a day for 21 days, followed by one seven-minute extinction trial a day for five days.[6] In addition, a control group of four males and four females was introduced to provide a base line for determining the relative strength of conditioning in the other two groups. These animals were familiarized with the situation by daily trials, but given neither reward nor finality. They were treated exactly like the other groups save that their training trials lasted one minute every day, no matter whether or not they pressed the bar or how often. The only exception was that they were not removed within 10 seconds of a pressure, to avoid giving the response the character of finality.

Results

Two measures of strength of conditioning were obtained: (1) 'latent time,' *i.e.*, the time it took the rat to press the bar after being put in the box; and (2) response frequency, or the number of responses made during extinction trials.

In Table 1 and Fig. 1 appear in condensed form the latent times

TABLE 1

Average Latent Times in Seconds of Bar-Pressing Response*

Experiments 1 and 2

Group	N	Training Trials							Extinction		
		1	2–4	5–7	8–10	11–13	14–16	17–19	1	2	3
Reward.....	12	370.9	155.9	46.6	31.6	20.1	20.9	19.2	16.7	25.0	16.9
Removal.....	12†	188.0	253.3	120.4	145.3	102.1	144.3	84.2	89.3	178.9	163.4
Control.....	8								108.6	220.1	192.3

* In Experiment 1, trials without a response were scored 900; in Experiment 2, training trials were scored 600, extinction trials 420.

† On the 20th day of training one male fell sick; extinction scores are therefore based on 11 cases in this group.

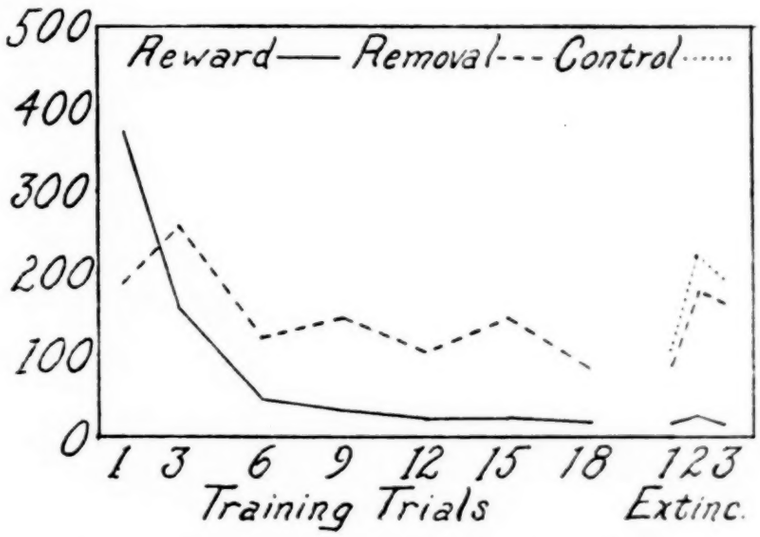

Fig. 1. Average latent time in seconds (ordinate) of bar-pressing response. Data from Table 1.

[6] In the training trials if five minutes elapsed without a pressure the experimenter started jiggling the bar sideways. After 10 minutes he pressed the bar with the rat's forefoot and removed the animal. The latter occurred only twice.

of both experiments for 19 days of training and for the first three days of extinction.[7] The training trials were averaged in successive three-day units. The first trial was kept separate, since no effect of reinforcement could possibly appear until the second. It is included merely to indicate any possible difference between the two groups in initial bar-pressing tendency. The results of the control group were averaged only for extinction trials, since there were often no responses during the one-minute training periods.

The latent times of the reward group yield a smooth, negatively accelerated learning curve. Those of the removal group also give evidence of learning, but much less and more variable. This finding was confirmed by analysis of individual scores. Comparison of the averages of days 2–7 and 14–19 showed clear improvement in 10 out of 12 cases in the reward group, 6 out of 12 in the removal group.

TABLE 2

AVERAGE FREQUENCY OF BAR-PRESSING RESPONSE DURING EXTINCTION. EXPERIMENT 1

Group	Reward (N =4)			Removal (N =4)		
Day	1	2	3	1	2	3
0– 5′	34.3	25.5	14.3	17.8	14.0	9.5
5′–10′	11.3	11.8	9.3	4.0	18.8	3.3
10′–15′	3.5	9.0	3.3	8.0	14.5	5.3
Total	49.1	46.3	26.9	29.8	47.3	18.1

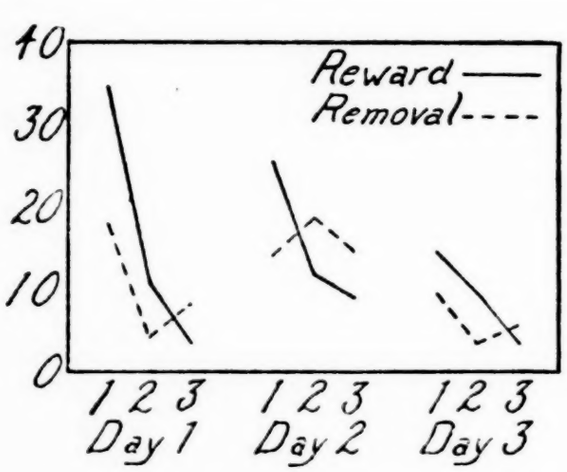

FIG. 2. Average frequency of bar-pressing response in successive five-minute periods of extinction. Experiment 1. Data from Table 2.

Owing to the difference in length of extinction periods the frequency of response data from the two experiments had to be treated separately. Experiment 1 is represented in Table 2 and Fig. 2.

[7] These were the limits within which data were available on all animals.

Here are shown, for the four rats of each group, the average response frequencies in successive five-minute periods.

The average number of responses per animal per trial for all three extinction days was for the reward group 41.7, for the removal group 33.3, an insignificant difference. More interesting is a comparison of the two groups period by period and day by day. The reward group shows the expected phenomena of extinction (**2**, **15**); *i.e.*, a pronounced drop within each trial, partial recovery between trials,

TABLE 3

AVERAGE FREQUENCY OF BAR-PRESSING RESPONSE DURING EXTINCTION. EXPERIMENT 2

Group	N	Day					Average
		1	2	3	4	5	
Reward	7*	48.0	45.7	22.6	13.9	20.7	30.2
Removal	7*	17.9	17.7	17.9	7.0	8.0	13.7
Control	8	9.9	2.9	8.4	3.6	5.0	6.0

* One male sick, results discarded.

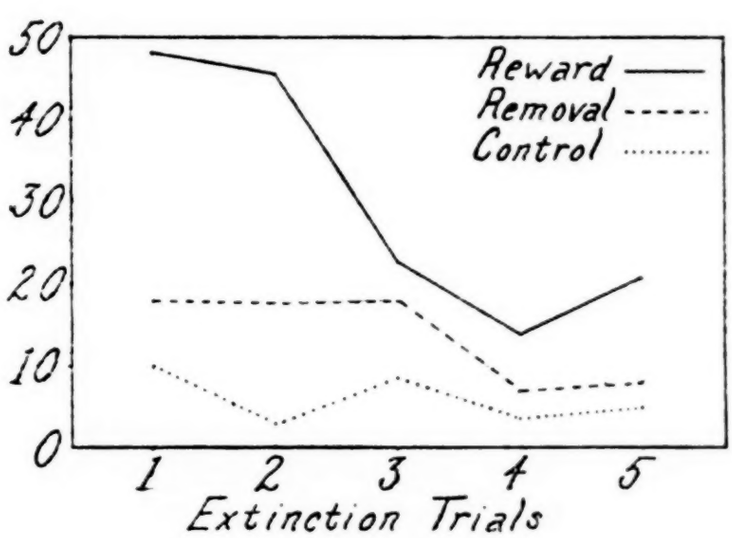

FIG. 3. Average frequency of bar-pressing response during extinction. Experiment 2. Data from Table 3.

and a decline from trial to trial. The removal group, however, shows no consistent trend, either within any trial or from day to day. It starts each trial below the reward group and finishes above it. The individual scores confirm the average trends. In the reward group nine out of twelve cases show a consistent drop from first to last period of a trial; in the removal group, five out of twelve.

Although the cases are too few to warrant a definite conclusion, the results suggest the operation of a factor in the reward group which is missing or, more probably, obscured in the other. Such a uniform

process would be covered by the concepts of inhibition and spontaneous recovery, or, possibly, of competition between two opposed sign-learnings. If the response tendency in the removal group were weaker it might appear more spasmodically and hence fail to show the same regularity of extinction. The same might be expected if, as seems likely, the response tendency in this group was aroused by some external rather than internal stimulus.

The response frequencies of the three groups in Experiment 2 appear in Table 3 and Fig. 3. Only the totals for the seven-minute periods are presented, since inspection of the scores minute by minute revealed no definite trend. In harmony with the earlier experiment, the reward group shows both a higher initial frequency and an earlier, steeper fall than the removal group. The latter in turn bears a somewhat similar relationship to the control. The averages in the final column of Table 3 were found to be reliably distinct. Both the reward-removal difference, with a *t*-value of 4.04, and the removal-control difference, with a *t*-value of 3.58, have probabilities through random sampling of less than .01.

It may be noted that the reward groups' scores for the first three days in Experiments 1 and 2 (Tables 2 and 3) are in fairly close agreement. The results of the two experiments could be combined, however, only for the first five minutes of the first day. When this was done it yielded average frequencies as follows: Reward (N = 12), 40.8; Removal (N = 11), 15.6; Control (N = 8), 4.9.

Discussion

We may now examine Guthrie's theory in the light of these results. First, is a response reinforced by its own finality? If we assume that the removal group represents the effect of finality alone, the answer is 'yes.' Latent time decreased in half the cases and for the group as a whole, and the number of responses during extinction was reliably greater than that of a control group in which responses had not been immediately followed by removal. It may be objected that the two groups differed not only in finality but also in frequency of response during training. This is true; the removal group actually averaged 19.4 responses to the control group's 16.4. Although it is regrettable that this factor was not controlled, it is unlikely that it was sufficient to account for the results. So far, then, the finality hypothesis is confirmed.

Secondly, is the reinforcing effect of a reward limited to that of the finality it produces? If this were true the results of our reward group should conform closely to those of the removal. The answer, then, is definitely 'no.' The superiority of the rewarded animals was

marked, whether measured by latent time or by response frequency in extinction. In this respect Guthrie's theory appears inadequate.

A theory of learning, to account for the present results, must answer two questions: First, how did immediate removal of the animal strengthen the bar-pressing tendency? Secondly, how did the reward of a food pellet strengthen it so much more? The first theory to be considered is that of Guthrie himself. Although the finality hypothesis alone fails to answer the second question, his system as a whole contains further possibilities. One is his concept of 'emotional reinforcement' (**5**); excitement increases proprioceptive stimulation, thus providing added cues for arousing the response. "These added stimuli may become conditioners of movement and hence explain the increased certainty of conditioning (p. 105)." In the present experiment the rewarded rats gave the impression of greater excitement by the quick, tense movements with which they investigated bar and food pan. Some objective indication of this might be obtained by measuring the intensity of pressure on the bar. The intensity of a response may itself be a factor in its rate of conditioning.

A second possibility within the framework of Guthrie's system is suggested by a recent experiment of Wolfe and Kaplon (**14**). They found that the chickens which were allowed the most consummative activity (pecks) showed the smallest variety of acts during the solution of a simple detour problem. In our experiment the rewarded rats limited their activities much more closely to the food pan and bar than did the others. According to Guthrie (**5**), the less varied the activity the fewer the *new* stimuli to be associated with the response on each occasion; therefore, the more rapid the learning. But the limiting of activity by consummative behavior remains to be explained.

The second type of theory to be considered is that which stresses the law of effect. The chief difficulty here is to explain, not the superiority of reward, but the efficacy of removal. Being lifted out of the box could hardly have the significance of escape, since the rat was transferred to a strange cage, in most cases as small as the experimental box.[8] 'Effect,' however, may take the form of delayed reward. In this study the delay varied from 15 to 30 minutes in the first experiment and from 30 to 75 minutes in the second, before the rats were returned to the living cages and fed. Few attempts have been made to measure the effect of delays of this magnitude on learning. In Wolfe's (**13**) study groups of rats which were delayed 10 and 20 minutes after the choice point in a T-maze made 58.5 percent and 54.8 percent correct choices, respectively. In Hamilton's

[8] No difference was noted between the results in Experiments 1 and 2 which could be attributed either to the introduction of the small detention box or to the longer detention period.

(6) maze-learning experiment the longest delay was seven minutes, but it is noteworthy that this group was no more inferior to the control group in trials and errors than were those with shorter delays. It is conceivable that in these studies, as in ours, the detention compartments became subgoals through association with ultimate reward and thus acquired reinforcing capacity.

Despite the plausibility of the foregoing theories, I should like to suggest another interpretation, based on observations of behavior in the experimental situation. I am quite convinced that almost all of the animals in both the reward and removal groups learned something over and above bar pressing. The two groups differed primarily not in amount of learning but in what they learned. This was indicated by their behavior in the vicinity of the bar. The typical rewarded rat after the first trial or two promptly nosed the food pan and hole, returning to them again and again. More gradually the bar became an object of attention, but always as an integral part of food-seeking activity. The removed rats were more variable in their behavior. However, eight of the ten on whom observations were recorded came to anticipate removal. They approached the bar, pressed it with a forefoot, quickly swung around, sometimes looking up toward the experimenter, and waited motionless a second or two. That this was not merely a reaction to the click of the signal magnet or the shadow of the experimenter's descending hand was proved by the many cases in which it occurred after the bar had been touched too lightly to record. Moreover the rats adopted different attitudes toward the bar. Some pressed with little delay and held themselves up toward the experimenter. Others displayed conflict behavior, hesitating before the bar, sniffing or barely touching it and drawing back as if alarmed. There is no question that the bar had acquired significance.

How does this account for the different amounts of reinforcement of the bar-pressing response in the two groups? Let us assume that an association has been formed between the visual-proprioceptive complex of bar pressing and a subsequent response; *i.e.*, either eating or adjusting to removal. The experimental box, and especially the bar, arouse incipiently both of the associated reactions; pressing-eating in one case, pressing-adjusting in the other. How quickly or frequently the bar is pressed, if at all, will then depend primarily on the strength of the tendency to complete the associated reaction. In a hungry animal the impulse to eat easily develops a powerful 'potential' to irradiate into the already aroused bar-pressing mechanism. There is presumably no similar drive leading to the behavior appropriate to being removed. Some rats, as noted, alternated between avoidance and approach behavior. On the whole,

however, the partial arousal of this response exerts an excitatory effect. Actually the two groups do not fit equally well into such a pattern. In the reward group food seeking appeared first; only later could we say that bar pressing was associated with it. In both groups, however, it is not a distortion to say that pressing the bar was reinforced only in so far as it became the antecedent to a consequent response.

Admittedly this is in no sense a formal theory of learning, though it may contain the germ of one.[9] Without going into its underlying assumptions, we may note one deduction which would provide verification apart from Guthrie's theory. The present argument attributes learning through removal, not to finality, but to the arousal of an anticipatory response to removal itself. The implication is that any stimulus sufficiently intense to arouse a definite reaction (so long as it is not a negative one), if applied repeatedly after a response, will reinforce that response to some degree. Thus, a noise of moderate intensity, turning a light on or off, or squeezing the animal momentarily each time it pressed the bar should to some extent decrease the latency and increase the frequency of the response.[10] A crucial test of Guthrie's theory, on the other hand, would involve removing the stimulus without evoking a characteristic response. Perhaps removing the bar instead of the rat would come closer to filling this requirement.

Summary

An experiment was performed to test Guthrie's theory that the function of reward is to remove stimuli and thus prevent unlearning of the final response. One group of rats was given a food reward upon pressing a bar; another group was simply removed from the experimental situation.

Both groups learned to press the bar, as measured by latent time and by frequency of response during extinction; the reward group, however, was clearly superior to the removal group by both measures.

Implications of these findings for theories of reinforcement are discussed and a hypothesis proposed.

(Manuscript received July 21, 1941)

[9] The reader will recognize a relation to White's (12) completion hypothesis of reinforcement and to Mowrer's (10) conception of the role of preparatory set. He will also recognize Hilgard's and Marquis's (7) threefold classification of theories of reinforcement. Although the account here advocated emphasizes 'expectancy,' I believe that both this and 'effect' can eventually be described in terms of 'substitution.'

It was only after the manuscript was submitted that I recognized the similarity of the present account to Tolman's law of emphasis (11, p. 387). The attempt is here made, however, to explain the strengthening of a *response* rather than that of a sign—gestalt.

[10] As suggested by Dr. N. E. Miller in conversation.

References

1. Elliott, M. H., and Treat, W. C., Hunger-contractions and rate of conditioning, *Proc. Nat. Acad. Sci.*, Wash., 1935, **21**, 514–516.
2. Ellson, D. G., Successive extinctions of a bar-pressing response in rats, *J. gen. Psychol.*, 1940, **23**, 283–288.
3. Guthrie, E. R., The effect of outcome on learning, *Psychol. Rev.*, 1939, **46**, 480–484.
4. Guthrie, E. R., *The psychology of human conflict*, N. Y.: Harpers, 1938.
5. Guthrie, E. R., *The psychology of learning*, N. Y.: Harpers, 1935.
6. Hamilton, E. L., The effect of delayed incentive on the hunger drive in the white rat, *Genet. Psychol. Monog.*, 1929, **5**, 131–207.
7. Hilgard, E. R., and Marquis, D. G., *Conditioning and learning*, N. Y.: Appleton-Century, 1940.
8. Holt, E. B., *Animal drive and the learning process*, N. Y.: Holt, 1931.
9. McCulloch, T. L., The role of clasping activity in adaptive behavior of the infant chimpanzee: III. The mechanism of reinforcement, *J. Psychol.*, 1939, **7**, 305–316.
10. Mowrer, O. H., Preparatory set (expectancy)—a determinant in motivation and learning, *Psychol. Rev.*, 1938, **45**, 62–91.
11. Tolman, E. C., *Purposive behavior in animals and men*, N. Y.: Appleton-Century, 1932.
12. White, R. K., The completion hypothesis and reinforcement, *Psychol. Rev.*, 1936, **43**, 396–404.
13. Wolfe, J. B., The effect of delayed reward upon learning in the white rat, *J. comp. Psychol.*, 1934, **17**, 1–21.
14. Wolfe, J. B., and Kaplon, M. D., Effect of amount of reward and consummative activity on learning in chickens, *J. comp. Psychol.*, 1941, **31**, 353–361.
15. Youtz, R. E. P., Reinforcement, extinction, and spontaneous recovery in a non-Pavlovian reaction, *J. exp. Psychol.*, 1938, **22**, 305–318.

TIME-ERRORS IN JUDGMENTS OF VISUAL EXTENTS

BY P. V. MARCHETTI

Rutgers University

Ever since Fechner (**3**) it has been well known that in the field of sensory intensities a symmetrical series of comparison stimuli, if presented more than 3 seconds after the standard stimulus, yield a marked preponderance of judgments in the direction of increase, *i.e.*, heavier, louder, brighter, as the case may be. At the time the judgments are made the standard stimulus is over and gone. It may therefore be assumed that its place has been taken by some physiological after-effect or trace. A preponderance of greater judgments is generally referred to as a negative time-error, in accordance with the terminology first used by Fechner; a preponderance of lesser judgments is called a positive time-error.

The phenomenon of the negative time-error has acquired renewed interest because of the treatment given to it by Köhler (**6**). The generality of its occurrence has led to the view that fading or sinking is one of the formal properties of neural traces (**6**, **10**). Not only do sensory intensities show the effect of the time-error, but sensory qualities, durations, affective processes, perceptive organizations, and even certain higher mental functions show some negative time-error when subjected to judgment in temporal sequence.

Danzfuss (**2**) presented pairs of chords to his subjects and asked them to judge the second as more or less agreeable than the first. His results showed a negative time-error, for the second was most frequently judged the more agreeable. Beebe-Center (**1**) obtained similar results with tones and odors. Wada (**12**) has shown negative time-errors in pitch; and Fodor and Happisch (**4**) also report them in judgments of taste and smell.

In view of the significance of the time-error as an indication of what happens to the mnemonic trace with the passage of time, it would seem of importance to know whether any sensory phenomena fail to reveal the negative time-error. But since the negative time-error represents the rule in all ordinary judgments of sensory intensities and frequently in other sensory and perceptual dimensions, as well as in affective judgments, it has been generally assumed that all psychophysical judgments under ordinary circumstances show this phenomenon with an interval between the critical stimuli greater than 3 seconds. Recent experiments, however, indicate that judg-

ments of bare visual extents show a positive time-error with intervals as large as 7 to 12 seconds.

The original interest of the author was to study the mnemonic traces of certain geometric illusions. It was soon found that more would have to be known of the fate of the traces of bare extents. After the first indication of a positive time-error, where the negative one had been the natural assumption, the entire experimentation was devoted to the study of this unexpected result.

Experimental Procedures and Results

There were six experiments employing the method of constant stimuli. The following instructions were given the observers in these experiments.

> "Your will be shown a black line on a white cardboard. After a brief interval another line will be shown to you. You are asked to judge the length of the second line in terms of the length of the first, *i.e.*, whether the second is *equal* to, or *shorter* or *longer* than the first. Make your judgment as soon as you can after you see the second line."

In the different experiments there were different intervals and different lengths of lines, but all gave a definite indication of the positive time-error. A typical experiment was one done with 5 observers wholly inexperienced in psychophysical judgments and completely ignorant of the phenomena of time-errors. The stimuli were black lines about 1 mm wide drawn on white cards 13 inches square. The standard was 100 mm long and the comparison stimuli were respectively 96, 98, 100, 102, and 104 mm long. These stimuli were presented tachistoscopically. Quite similar results, however, have been obtained with cards as large as 28 × 44 inches held up by the experimenter in front of a group of observers. With the larger cards the lines were even as long as 30 inches. Stimuli as long as 5 and 8 feet have been used.

The interval between the critical stimuli with the tachistoscopically presented material was 8 seconds. The standard was, of course, shown before each variable. The variables were presented in haphazard order determined before the experiment by drawing from a scrambled lot of numbered slips of paper. Fifty judgments were made on each variable by each of the 5 observers.

The results are given in Table I. The columns headed *V* are the results obtained with vertical lines and *H* designates the results

TABLE I

S	*IP*		*PSE*	
	V	*H*	*V*	*H*
1	10.0	32.0	100.50 mm	101.60 mm
2	−1.6	22.8	99.42 mm	101.14 mm
3	7.2	12.8	100.36 mm	100.66 mm
4	19.6	34.0	100.98 mm	101.70 mm
5	10.8	19.0	100.54 mm	100.96 mm
Av.	9.2	24.1	100.36 mm	101.21 mm

obtained with horizontal lines. *IP* is the index of preponderance, the formula for which is

$$IP = \frac{S - L}{N}(100),$$

where S is the number of judgments of shorter, L is the number of judgments of longer, and N is the total number of judgments made. A positive IP indicates a positive time-error and a negative one a negative time-error. The point of subjective equality (PSE) was derived by the summation method (**13**).

Although there is one small negative time-error for the second subject in his judgments of vertical lines, the experiment as a whole demonstrates a decidedly positive time-error in the judgments of bare visual extents, both vertical and horizontal, even with an interval as great as 8 seconds. The time-error appears to be greater for the horizontal lines than for the vertical.

The effect of background on the apparent length of the lines was studied by varying the size of the cards on which the standard and comparison stimuli were shown. The results were what would be expected, *i.e.*, if two lines are of equal length, but one is drawn on a smaller card than the other, the one on the smaller card appears longer. By placing the comparison stimuli on smaller cards than the standard, it is possible to produce a negative time-error. When they are on cards larger than the standard, the time-errors are markedly positive. But when the standard and comparison cards are the same size, only the positive time-error appears.

The phenomenon of assimilation (**7**) was also studied by the interpolation of a stimulus between the two critical stimuli. The results show that traces of bare visual extents readily assimilate to one another. With the interpolation of a stimulus much shorter than the standard there is a definite negative time-error; with the interpolation of a stimulus greater than the standard, there is an increased positive time-error.

It was thought that the persistent positive time-errors might in some way be related to the phenomenon of the 'central tendency' as described by Ipsen (**5**), or to Leuba's (**8**) 'law of sense memory.' To study this possibility the method of single stimuli was employed with two different series of stimuli relatively far apart, *i.e.*, one series of short lines and one of long lines. The results indicated a positive time-error for both the short lines and the long lines. The time-errors for short (10 cm) and long (8 ft) lines were not sufficiently different to allow the assumption that the time-error varies with the length of line judged.

A simple demonstration of the positive time-error was made with a class of twenty students. A 30-inch line was drawn on the blackboard, and the students were asked to note its length in order to compare it with a line to be drawn on the board the next day. Twenty-four hours later another line again thirty inches long was drawn on the board. Sixteen of the 20 members of the class judged

this second line to be shorter than the one presented the day before. This result reveals a large positive time-error (*IP* 60) for the group. It also tends to reinforce the view that psychophysical time-errors, since at least in this case they operate over a period of twenty-four hours, are genuine mnemonic processes.

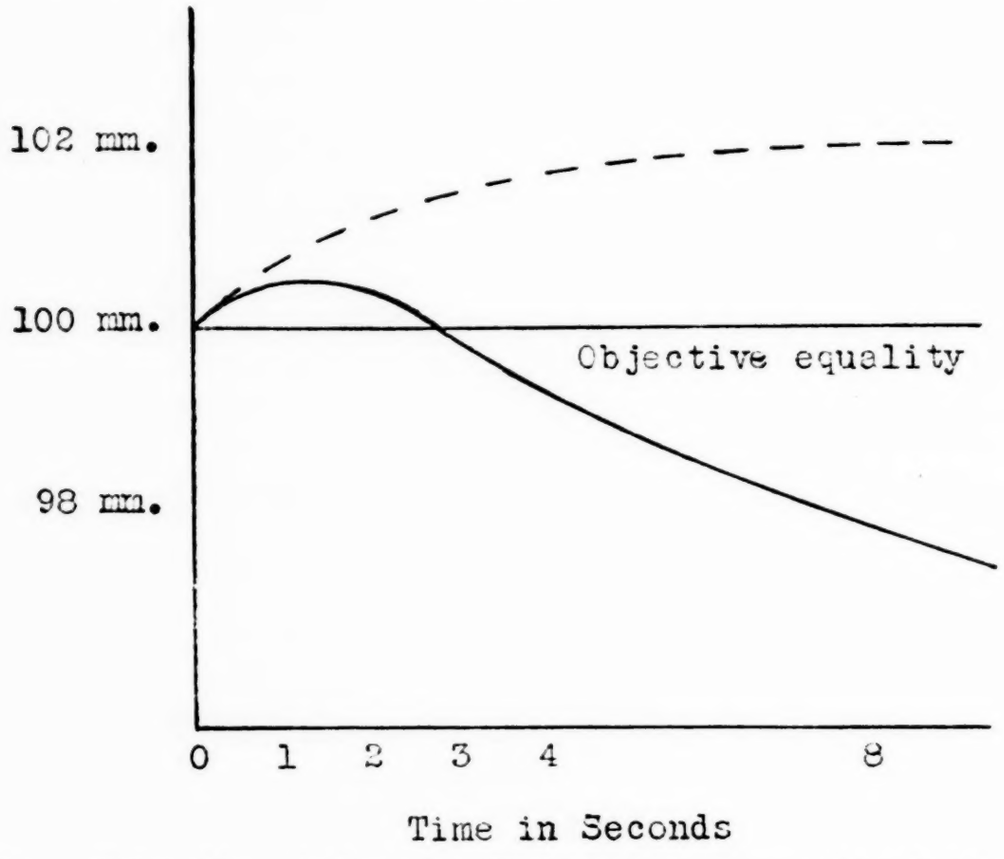

FIG. 1. Traces of bare visual extents. The broken curve represents hypothetical values of traces of visual extents. The *PSE* used for plotting the curve is the average for both horizontal and vertical lines at the end of the interval of 8 seconds. The solid curve is the one drawn by Köhler (6, 159) to represent the course of traces of sensory intensities.

SUMMARY

1. Judgments of sensory and perceptual phenomena generally reveal a negative time-error whenever the stimuli are separated by intervals longer than three seconds. The rule may possibly be extended to include all judgments that can be formulated in quantitative terms—heavier, brighter, louder, larger, better, nicer, *etc.* The second of two stimuli judged in such terms calls forth a preponderance of judgments in the direction of increase.

2. An exception to this rule appears to occur in judgments of bare visual extents, at least under the conditions reported in this experiment. Instead of a preponderance of judgments of longer, there is a preponderance of judgments of shorter in the case of black lines drawn on white cardboards.

3. The apparent length of a line is partially a function of the size of the background on which it is seen. The larger the background, the shorter the apparent length of the line, and *vice versa*. But when the size of the backgrounds of the standard and comparison stimuli is the same, the positive time-error still persists.

4. Judgments of visual extents reveal the operation of some factor like assimilation. Interpolation of a long line between the standard and comparison stimuli produces an even more pronounced positive time-error. Interpolation of a short line produces a negative time-error.

5. Within the range of stimuli used in this experiment there was no reliable evidence of a central tendency of judgment.

6. The time-error persists over a period of twenty-four hours, and may therefore be regarded as a phenomenon of memory.

(Manuscript received June 26, 1941)

References

1. Beebe-Center, J. G., *The psychology of pleasantness and unpleasantness*, New York: Van Nostrand, 1932, 197.
2. Danzfuss, K., Die Gefühlsbetönung einiger unanalysierten Zweiklänge, Zweitonfolgen, Akkorde und Akkordfolgen bei Erwachsenen und Kindern, *Phil. u. Psychol. Arbeiten*, 1923, **5**, 18 f.
3. Fechner, G. T., *Elemente der Psychophysik*, Leipzig: Breitkopf and Härtel, 1907 edit., vol. I, 341, vol. II, 560.
4. Fodor, K., and Happisch, L., Die Bedeutung der Zeit zwischen zwei Vergleichsreizen bei Bestimmung von Unterschiedsschwelle. I. Untersuchungen am Geschmacksorgan, *Arch. f. d. ges. Physiol.*, 1922, **197**, 337.
5. Ipsen, G., Ueber Gestaltauffassung, *Neue Psychol. Studien*, 1926, **1**, 171–278.
6. Köhler, W., Zur Theorie des Sukzessivvergleichs und der Zeitfehler, *Psychol. Forsch.*, 1923, **17**, 130–177.
7. Lauenstein, O., Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler, *Psychol. Forsch.*, 1932, **17**, 130–177.
8. Leuba, J. H., A new instrument for Weber's law, with indications of a law of sense memory, *Amer. J. Psychol.*, 1892, **5**, 370–384.
9. Needham, J. G., The time-error in comparison judgments, *Psychol. Bull.*, 1934, **31**, 229–243.
10. Pratt, C. C., The law of disuse, *Psychol. Rev.*, 1936, **43**, 83–93.
11. Pratt, C. C., Interaction across modalities: I. Successive stimulation, *J. Psychol.*, 1936, **2**, 287–294.
12. Wada, Y., Der Zeitfehler beim Sukzessivvergleich der Tonhöhen, *Jap. J. Psychol.*, 1932, **7**, 505.
13. Woodworth, R. S., *Experimental psychology*, New York: Holt, 1938, 421.

THE SIGNIFICANCE OF AUDIBLE ONSET AS A CUE FOR SOUND LOCALIZATION

BY DAVID P. BODER AND IRVING L. GOLDMAN

Psychology Laboratory of the Illinois Institute of Technology

It has been established by a number of investigators that distinctness of onset and distinctness of termination of a tone facilitate the localization of the source. The bibliography of the subject up to 1938 is well covered by Stevens and Davis (4) and by Woodworth (9). Valuable bibliographical data are given by Fernberger *et al.* (3) in their 1941 review of the topic. An article by Wilska (8) seems especially complete and is appended by a comprehensive bibliography.

The overwhelming number of experiments on sound localization was dominated by what we may call the 'sound cage' principle. In such experiments the variation of the position of the sound source occurs either by changing the distance while the angular position is kept constant (near-far variation); or by changing the angular position with the distance constant (lateral variation). Most of the experiments were performed indoors. Outdoor experiments are reported by Stevens and Newman (5) and Marx and Marx (3).[1]

The purpose of the present experiment was to approach the problem of sound localization under the following conditions:

(*a*) To devise a method for group testing under readily duplicable conditions.

(*b*) To determine the localization value of clear onset of tones *vs.* masked onset of tones of the same duration.

(*c*) To compare the localization value of prolonged tones of either clear or masked onset with the localization value of an interrupted tone. Such an interrupted tone must be considered as a sequence of stimuli even though a single response is required from the observer.

(*d*) To compare the responses under indoor and outdoor conditions.

(*e*) To present the *O*s with a constellation of stimuli which vary in position not along a circle, the center of which is within the head of the observer, but along the base of a triangle the vertex of which is

[1] This article by Marx and Marx is the only one found by the present writers which deals directly with the psychological problems of sound localization under military conditions. It reports in a somewhat sketchy manner the procedure of determining the positions of enemy batteries in the Argonne during a period of 18 months. A footnote further indicates that by the end of the World War No. 1 the German army possessed special military detachments dealing with sound-finding problems.

located on the aural axis[2] of the observer. In other words, the sound sources were placed on a horizontal line perpendicular to the line of sight of the observer; they were before him as objects hung on a straight wall or fence.

(*f*) In all experiments the *O*s had their eyes open and had complete freedom of motion; the only posture restriction implied in the instructions was that the *O*s remain seated on their chairs.

Apparatus

The apparatus (Fig. 1) permitted three kinds of stimuli to be produced: (*a*) a sequence of short tones with a definite time interval, *i.e.*, the intermittent stimulus (Intermt.); (*b*) a continuous tone with a definite onset or start, *i.e.*, the discrete-continuous (Disc.); (*c*) a continuous tone overlapping its predecessor, *i.e.*, the continuous-interlocking stimulus (Cont.).[3]

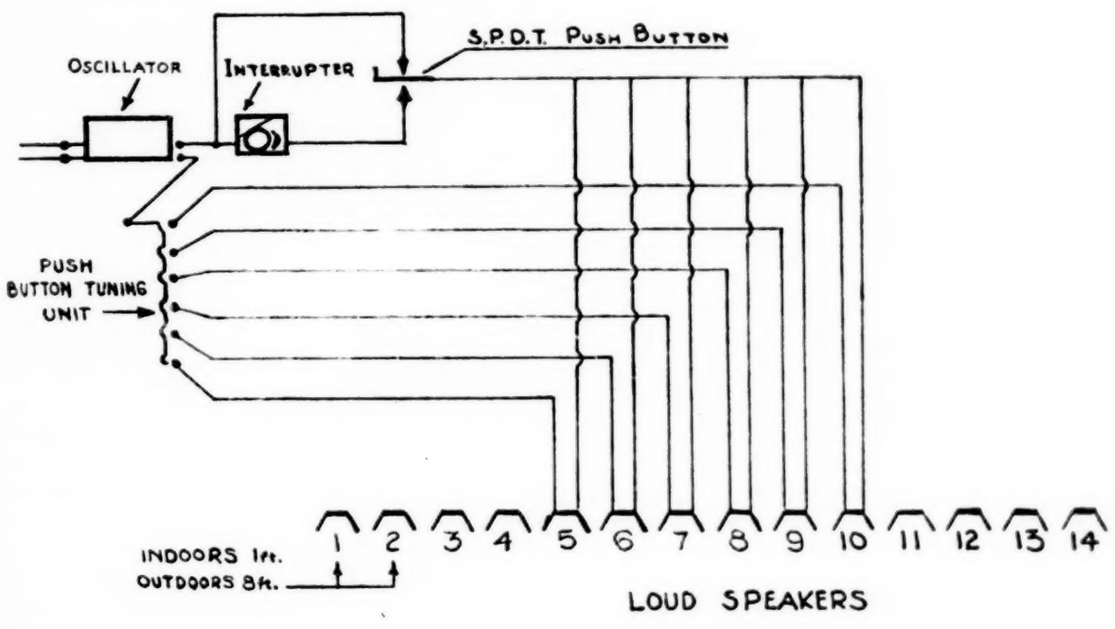

Fig. 1.

The apparatus consisted of the following parts:

(1) Six 3″ speakers for the indoor experiment and six 12″ speakers for the outdoor experiment.

(2) An audio oscillator, and an additional 25-watt amplifier for the outdoor experiment.

(3) A push-button switch with six positions, each corresponding to a definite speaker.

(4) A telechron driven interrupter consisting of a disk of insulating material with a metal contact of about 30°. The disk revolved at the rate of one revolution per sec. and the ratio of the duration of the tone to the silent interval was 1 : 11.

(5) A single pole, double throw switch for the conversion of continuous stimuli into intermittent stimuli.

(6) Eight 12″ speaker cones to simulate speakers for the inactive positions in the outdoor experiment.

---

[2] By aural axis is meant the straight line connecting the observer's two ears within his head.

[3] Mr. John O'Malley has assisted in the assembly of the equipment for the preliminary experiments.

Procedure

The *O*s were given mimeographed answer sheets and were instructed that only one speaker would sound at a time. They were to listen carefully and when the examiner called out a letter, they were to place the number of the sounding speaker next to the letter on the answer sheet. The experimenter would call the letter about two seconds after the assistant had started the stimulus. Each stimulus lasted 15 seconds. Each experiment consisted of four preliminary trials, and three sections of twelve trials each. The *O*s were not told whether their answers in the preliminary trials were correct or not. The preliminary exercises were simply given to familiarize the *O*s with the situation. Each section of twelve stimuli contained four stimuli of each type given in a mixed but prearranged order.[4] The sequences of stimuli were different in each section. At the end of each section the subjects were instructed to fold the answer sheet along a printed horizontal line so that the completed sections were covered from sight during subsequent work. This was done to avoid a possible attempt on the part of some observers to guess the answers by avoiding repetition of sequences.

Experiment I

This series was given indoors. A partition or screen about 16 ft wide and 6.5 ft high was placed in the front of the *O*s. Fourteen 1″ holes, numbered correspondingly, were drilled in the screen at the level of about 45″ from the floor, and one ft apart (measured from center to center of the holes); this amounts to an angular distance between the holes of about 2.85°. Behind the middle 6 holes, from No. 5 to No. 10, were placed six 3″ magnetic speakers. All holes were covered from behind with thin opaque cotton material so that the *O*s were unaware that the extreme four holes on either side had no speakers behind them. This was done in order to avoid an undue number of "chance correct" answers at the two extremes (tail effect). Under such a setup the extreme speakers permitted a stimulus error of nine steps to one side and only four steps to the other side. (The experimental results (Tables II and IV) seem to indicate that at least indoors, the number of non-active speakers at the extremes should have been still greater, since about 24 percent of the errors for the Cont. stimulus exceed 4 steps.)

The subjects were seated for this experiment in a class room, the nearest row 10 ft and the farthest row 30 ft away from the screen. The spread in width of the seats was slightly less than the distance between speaker No. 2 and No. 13, *i.e.*, 10 ft. With proper separation there were 3 persons in a row. No attempt was made to soundproof the room or to eliminate ordinary noise. The experiment was performed in the laboratory on the fourth floor of the Madison Street building in Chicago, on an intersection of busy trolley-car lines. Windows were closed. The noise level[5] in the room without the speaker sounding was fluctuating between 55 to 60 db at the first row

[4] The orders of stimuli and speakers for both series, indoors and outdoors, are given in the key below. (Capital letters indicate place in answer sheet; *i*, *d*, *c* indicate types of stimuli: Intermt., Disc., Cont., respectively.)

Prelim. Series	O.	P.	Q.	Z.
Stimulus Type	*i*	*d*	*c*	*c*
No. of Speaker	5	10	6	9

First Series	A	B	C	D	E	F	G	H	I	J	K	L
Stimulus Type	*d*	*c*	*i*	*d*	*c*	*c*	*i*	*d*	*c*	*i*	*i*	*d*
No. of Speaker	5	8	10	7	5	9	8	10	7	5	9	6

Second Series	A	B	C	D	E	F	G	H	I	J	K	L
Stimulus Type	*i*	*d*	*c*	*i*	*d*	*i*	*d*	*c*	*i*	*d*	*c*	*c*
No. of Speaker	8	10	5	9	6	10	8	6	7	9	6	9

Third Series	A	B	C	D	E	F	G	H	I	J	K	L
Stimulus Type	*d*	*c*	*i*	*i*	*d*	*c*	*d*	*i*	*d*	*c*	*c*	*i*
No. of Speaker	8	10	5	7	5	8	9	6	7	10	7	6

[5] The authors acknowledge with thanks the cooperation of the Armour Research Foundation and of Messrs. M. Camras and L. T. Anderson for assistance in the sound level measurements.

of seats (nearest to the street), and 50 to 60 db at the last row (near a wall having windows to an alley). With one speaker sounding, the noise level was from 64 to 70 at the first row and 58 to 64 db at the last row. The intensity as well as the pitch of the oscillator also varied slightly during a 15 minute session. No increase above the noise level in the room was detected by the measuring instrument at the last row during the Intermt. stimuli. The latter was however clearly registered at the first row by a jerk of the indicator of the sound level meter of about 3 db.

The results were scored in terms of steps of error from the correct speaker; so, if a speaker was named correctly, the score for the answer was *zero;* if the nearest speaker to the right or left of the correct one was given, the response was rated as one step error and the score was *one;* if the answer was two speakers away from the correct one, the score was *two*, etc. The error scores for each type of answers of a given individual were averaged separately, so that his final rating was given in terms of three error scores for the corresponding three types of stimuli. The results for this group are given in Tables I and II. No individual differences which could be attributed to variation of distance from the stimulus source within the used seating arrangement could be detected.

TABLE I

INDOOR SERIES

Summary of the Results of the Indoor Experiment in Terms of Error Scores ($N = 95$)

Type of Stimulus	Range of Error Scores in Steps (one ft. per step)	Av.	SD	$\sigma_{av.}$
Intermt.	0.17 to 5.83	2.11	1.07	.11
Disc.	0.58 to 5.42	2.60	1.39	.14
Cont.	1.50 to 4.75	2.98	1.21	.12

	Cont. and Disc.	Intermt. and Disc.	Intermt. and Cont.
Difference (in 1 ft. steps)	0.38	0.49	0.87
Critical Ratio	2.06	2.72	5.37

TABLE II

SUMMARY OF RESULTS OF THE INDOOR EXPERIMENT IN TERMS OF FREQUENCY OF CORRECT ANSWERS AND FREQUENCY OF STEPS OF ERROR MADE BY THE GROUP

	Intermt.		Disc.		Cont.	
	N	%	*N*	%	*N*	%
Correct answers	201	17.6	155	13.6	117	10.3
Errors of						
1 Step	330	28.9	266	23.3	227	19.9
2 Steps	220	19.3	196	17.2	184	16.1
3 Steps	133	11.7	182	16.0	181	15.9
4 Steps	131	11.5	131	11.5	150	13.1
5 Steps	62	5.4	82	7.2	125	10.9
6 Steps	25	2.2	71	6.2	71	6.2
7 Steps	14	1.2	26	2.3	47	4.1
8 Steps	15	1.3	22	1.9	29	2.5
9 Steps	6	0.5	9	0.8	8	0.7
Answers omitted	3	0.3	0	0	1	0.1
Total	1140		1140		1140	

Experiment II

The outdoor experiment was undertaken on the athletic field of the Illinois Institute of Technology. This is located in a section of empty lots and low buildings for several blocks, except for the southern border where the four- and five-story buildings of the Institute are located. Again fourteen positions were used of which unknowingly to the *O*s only the middle six consisted of speakers connected to the oscillator. These were 12″ P.M. Utah speakers,[6] fastened to a fence at the eastern boundary of the field at about 6 ft 5 in. from the ground to the center of the speaker. The distance between the speakers from center to center was 8 ft corresponding to an angular distance of about 2.25°. At a distance of 200 ft from this fence seven rows of chairs were located, 9 in a row, the center chair corresponding to the mid-point between speaker No. 7 and No. 8. Only five *O*s were seated (on alternate chairs) in each row.

TABLE III

Outdoor Series

Summary of Results of the Outdoor Experiment in Terms of Error Scores (N = 88)

Type of Stimulus	Range of Error Scores in Steps (8 ft per step)	N = 88		
		Av.	SD	$\sigma_{av.}$
Intermt.	0.4 to 2.8	1.33	0.50	.05
Disc.	0.7 to 3.4	1.75	0.64	.07
Cont.	1.3 to 3.7	2.08	0.63	.07

	Cont. and Disc.	Intermt. and Disc.	Intermt. and Cont.
Difference (in 8 ft. steps)	0.33	0.42	0.75
Critical Ratio	3.3	4.8	8.7

TABLE IV

Summary of the Results of the Outdoor Experiment in Terms of Frequency of Correct Answers and Frequency of Steps of Error Made by the Group

	Intermt.		Disc.		Cont.	
	N	%	N	%	N	%
Correct answers	283	26.8	219	20.7	166	15.7
Errors of						
1 Step	384	36.4	331	31.3	296	28.0
2 Steps	235	22.2	224	21.2	233	22.0
3 Steps	94	8.9	149	14.1	161	15.2
4 Steps	42	4.0	76	7.3	103	9.8
5 Steps	15	1.4	33	3.1	58	5.5
6 Steps	1	0.1	18	1.7	18	1.7
7 Steps	1	0.1	5	0.5	12	1.1
8 Steps	1	0.1	0	0.0	7	0.7
9 Steps	0	0.0	1	0.1	2	0.2
Total	1056		1056		1056	

[6] The twelve-inch speakers were Utah P.M. speakers with 4-lb magnets. Thanks are due to the Utah Radio Products Co., Chicago, for the loan of the speakers and to Messrs. P. Heckendorn and C. Walker, engineers of the firm, for valuable advice.

The noise level on the field with the speakers silent fluctuated between 70 and 80 db. When the speakers were *on*, the noise level increased up to about 82 at the first row and remained between 70 and 80 db at the last row. One item presented a serious difficulty. Passenger and freight trains passed several times during the experiment on tracks about 150 ft behind the observers' seats. In anticipation of this disturbance the *O*s were told not to write their answers when a train passed, and the stimulus was repeated after the noise of the trains had died down. This presented no difficulty with intermittent or discrete stimuli; if the interruption occurred at a continuous-interlocking stimulus a tone of another speaker was given for about three seconds preceding the repetition of the stimulus tone as such.

The experimenter in charge of the instructions and calling out of the letters used in the outdoor experiment a separate amplifying system with a speaker placed near to the chairs of the observers. He himself had to be about fifty feet from the assistant operating the stimulus device. Care had to be taken to shield the microphone of the experimenter so that it would not pick up the stimulus notes from the experimental speakers.

In this outdoor experiment the differences between the localization values of the three types of stimuli stand out with definite clearness. The data are summarized in Tables III and IV.

Discussion

Comparing Tables I and III we notice in the latter: (*a*) a narrower range of the scores, (*b*) the markedly lesser upper level of the ranges, (*c*) smaller SD's, and (*d*) larger critical ratios for all three pairs of differences, indicating definite reliability. We see that in the indoor series only the difference Intermt. and Cont. has a critical ratio better than 3, the other two critical ratios being below 3. Tables II and IV confirm the advantage of the outdoor situation in terms of significantly larger percentages of correct and nearly correct (*i.e.* one step errors) responses.

What were the factors contributing to such improvement of scores?

TABLE V

A Comparative Presentation of the Conditions of the Indoor and Outdoor Experiments

	Indoor Experiment	Outdoor Experiment
No. of *O*s	95	88
Educational level	College	College
Sex	Male—76; Female—19	All male
Size of Loud speakers	3″	12″
Distance between speakers	1 ft	8 ft
Distance of *O*s from speakers	10–30 ft	200–220 ft.
Angular dist. between speakers in degrees	about 2.85°	about 2.17°
Total sound level	58–70 db	76–82 db
Distractions	Few: windows closed	Street noises & passing trains
Height of speakers	45 in. from the floor	77 in. from the ground
Angle of speakers to horizontal plane of aural axis?[7]	about 0.2° downward	about 0.7° upward

[7] The average distance of the ears from the ground computed on 10 sitting adults amounted to about 46 inches. The slightly higher position of the speakers in the outdoor experiment was due to unforeseen technical reasons. In the light of the work of previous investigators, this variation in the vertical direction should be considered immaterial for the precision of exclusively lateral localization, which was the case in our investigation.

The best known physical factor favoring outdoor localization is the absence or near-absence of standing, secondary waves which are present indoors because of the reflection of the walls of the room. Among other factors we could consider the possible "constancy" attributes of the stimuli which would suggest that a more powerful speaker at a far-away distance could be more readily localized than a weak speaker at a proportionally nearer distance. A comparative presentation of the conditions in both experiments is given in Table V.

We now turn to the individual differences. James and Massey (2) have already indicated that some individuals localize continuous tones better than intermittent tones. The scientific as well as practical implications of this phenomenon, if consistent, are far-reaching. A person with such an 'inversion' of abilities may possibly be considered specially endowed as well as handicapped. We recently hear of attempts to make special use of color blind individuals for military purposes. It may be useful to single out for certain purposes those whose auditory localization is especially good for continuous tones.

TABLE VI

RESULTS OF BOTH EXPERIMENTS IN TERMS OF SCORE PROFILES OF THE OBSERVERS

Score Profile	Indoor $N = 95$		Outdoor $N = 88$		Total $N = 183$	
	N	%	N	%	N	%
1. Intermt. Disc. Cont.	40	42.1	42	47.9	82	44.8
2. Intermt. Cont. Disc.	18	18.9	17	19.4	35	19.2
3. Intermt. Disc.† Cont.	3	3.2	0	0	3	1.6
4. Intermt.† Disc. Cont.	4*	4.2	3*	3.4	7*	3.8
5. Intermt.† Cont. Disc.	1*	1.0	1*	1.1	2*	1.1
6. Disc. Intermt. Cont.	12	12.6	16	18.2	28	15.4
7. Disc. Cont. Intermt.	5	5.3	1	1.1	6	3.3
8. Disc. Cont.† Intermt.	1	1.0	0	0	1	0.6
9. Disc.† Intermt. Cont.	4*	4.2	3*	3.4	7*	3.8
10. Disc.† Cont. Intermt.	1*	1.0	0*	0	1*	0.6
11. Cont. Intermt. Disc.	5	5.2	6	6.8	11	6.0
12. Cont. Disc. Intermt.	5	5.2	1	1.1	6	3.3
13. Cont. Disc.† Intermt.	0	0	1	1.1	1	0.6
14. Cont.† Intermt. Disc.	1*	1.0	1*	1.1	2*	1.1
15. Cont.† Disc. Intermt.	1*	1.0	0*	0	1*	0.6

† A dagger between two stimuli indicates a tie in scores.
* Items marked with an asterisk indicate tied scores which appear twice in the table.

Table VI presents a summary of the individual records. The score profile indicates the order of proficiency from the best to the poorest with which the three types of stimuli have been localized by a given observer. As we notice, the percent of individuals of each category is highly similar, in the indoor and outdoor groups. This

similarity occurs in spite of the differences in the averages for the two situations. The best scores for localizations are obtained for the intermittent stimulus: 63.2 percent of the indoor series and 67.3 percent of the outdoor series corresponding to 65.6 percent of the total, or 120 individuals of the 183 participants in the two experiments (Table VI, entries 1, 2, 3). Only 35 individuals, or 19.3 percent of the group, obtained their best score in localizing discrete-continuous stimuli. The number of individuals who obtained their best score with the localization of the continuous-interlocking stimuli amounts to 18 persons, or about 9.9 percent of the group (entries 11, 12, 13). Score profiles with the intermittent stimulus ranking lowest occur only in 13 cases, or about 7.2 percent of the group (entries 7, 10, 12).

TABLE VII

INDIVIDUAL ANALYSIS OF RECORDS OF TYPICAL SCORE PROFILES GIVING THE ERROR SCORE (e.s.) AND QUARTILE (*i.e.*, q.4 THE LOWEST, q.1 THE HIGHEST QUARTILE)

Observers	Type of Experiment	Score and Quartile	Cont.	Disc.	Intermt.
(1) G. K.	Indoors	e.s.	1.50	1.83	2.08
		q.	1	1	2
(2) K. C.	"	e.s.	2.50	2.75	3.25
		q.	2	3	4
(3) M. P.	"	e.s.	3.17	5.32	5.43
		q.	3	4	4
(4) T. W.	"	e.s.	3.75	4.42	5.25
		q.	4	4	4
(5) K. K.	"	e.s.	3.92	4.25	4.75
		q.	4	4	4
(6) – G.	Outdoors	e.s.	1.5	2.00	2.10
		q.	1	3	4

The score profile Intermt., Disc., Cont. is obtained most frequently although not by the majority of the subjects. Only 42.1 percent of the *O*s of the indoor series and 47.9 percent of the outdoor series have such a profile. This would suggest that the discrete onset of a continuous tone lasting 15 seconds is by itself not enough of an advantage over a masked onset to produce better localization. Whether the advantage of the intermittent stimulus lies in the brevity of its components in addition to both clear onset and clear discontinuance cannot be determined from this series of experiments. The possibility is not excluded that the length of a continuous stimulus may obliterate the advantage gained by a clear onset. In our experiments a number of continuous-interlocking stimuli had a discrete discontinuance, but the observers were expected to write their answer before the stimulus ceased. A problem in psychophysics is here on hand.

The individual analysis of records (Table VII) shows that reversed score profiles occur in five indoor cases and one outdoor case, or in 5 percent and 1.1 percent respectively (Table VI, entry 12). One of the five indoor *O*s (G. K.) a left-handed woman, scored in all three types of responses in the upper half of the group; one observer (J. N.) gave a mixed score pattern and three have all scores in the lower half. The only outdoor observer (– G.) with a reversed score profile falls in the mixed group.

The consistency of such score profiles in these individuals could not be ascertained in the present project.

Two series of facts merit consideration in connection with these findings: one leads us in the direction of bio-electric phenomena; the other points in the direction of volitional behavior or *conative dynamics*, possibly related to Webb's *w* factor (**7**).

A. All investigators of action potentials in the auditory pathway agree that while the auditory nerve is apt to reproduce, within a wide range, the frequencies of a tone, the higher brain levels give, under present day conditions of amplification, only an indication of a change of potential at the onset and at the discontinuance of the tone. In this respect the higher auditory centers act similarly to the optic tract, where only *off* and *on* responses are obtainable (here possibly for no other reason than for the physical properties of light). It seems that the precision of visual localization is greatly enhanced by the interrupting devices provided by nature in the form of blinking, and the jerking fixation movements of the eye muscles. In other words the kinematic equipment of the visual apparatus is much more extensive and efficient than that of the auditory apparatus. One cannot help mentioning at this instance that birds which appear to possess a very keen gift of hearing are structurally fit for very fast and nearly semi-circular head movements. Dogs, whose efficiency in spotting the prey depends at least in part on auditory means, possess in addition to the pineal reflex the anatomical equipment, and behavior tendency of covering and uncovering of the auditory meatus with their ear flaps. This seems to permit the animal to superimpose a pseudo-intermittence or at least a pseudo-interruption over an otherwise continuous sound stimulus, such as embodied by the movements of the prey in the grass.

It seems, therefore, that for the higher perceptual integration of an auditory stimulus, certain bio-electrical properties, such as the *on-off* effect, may be of special significance.

B. In connection with the individual differences obtained within the same category of stimuli it appears that two groups of factors have to be considered. From the work of Wallach (**6**) and others it

is well established that the precision of sound localization may greatly depend not only on the position of the head but on the head movements as such. However, with the possible exception of very sudden or of very intensive sounds of the explosive type, auditory stimulation does not arouse such rigid reflex mechanisms as is the case of the eye and head movements in response to a moving object or a sudden flash of light in the peripheral field of vision. The head and body movements in response to sound appear much slower and normally more selective than is the case of vision. These movements are executed by a constellation of striped muscles most of which, at least in the human adult, have lost a great deal of their reflex-like character in favor of intentional selective motion patterns. Motivational and "intentional" factors seem to play in sound localization a much greater role than in visual space perception and from that standpoint the search for a 'personality factor' as a component of the process of auditory localization may be well worth while.

Conclusions

(1) It has been verified under 'everyday' conditions that *indoors* an intermittent tone of about 680 d.v. is localized better than a continuous tone, of the same frequency, with a masked onset.

(2) There is a tendency *indoors* for the intermittent tone to be localized better than a continuous tone with discrete onset, but the critical ratio of the difference between the respective error scores only approaches but does not attain the desired value of 3.

(3) The clearness of the onset of a continuous tone of the studied frequency is not sufficient *indoors* to render a reliable difference between the group averages of scores for continuous-interlocking and discrete-continuous stimuli, although the tendency is indicated by a critical ratio of the difference of slightly better than 2. We have a paradoxical situation akin to the addition of subliminal differences in psychophysics. Diff.$_{i.d.}$ is unreliable; diff.$_{d.c.}$ is also unreliable, but diff.$_{i.c.}$ becomes definitely reliable.

(4) *Outdoors* with the distance about 10 times greater than indoors, and in spite of the angular distance between speakers being somewhat less, all three differences between scores for the three types of stimuli were statistically reliable; the rank order of the averages for the three types of stimuli is outdoors the same as in the indoor experiments: intermittent best, discrete-continuous next, and continuous-interlocking last or poorest.

(5) Under so-called 'everyday' conditions the change from indoors to outdoors in sound localization experiments is bound to lead to changes in more than one variable and more than one attribute of the

stimulus (characteristics of the sound system, wind conditions, etc.). It seems therefore that for practical purposes each situation be definitely explored.

(6) The majority of observers rendered score profiles indicating that they rated best in the localization of the intermittent stimuli.

(7) The score profiles of individual observers (from best to poorest proficiency) take most frequently the score pattern of Intermt., Disc., and Cont.; this happens in 44.8 percent of the cases. In 19.2 percent of the cases the Intermt. is localized best but Cont. and Disc. had their ranks reversed to third and second respectively. Cont. appears as best localized only in 18 records, or in 9.9 percent of the cases. Completely reversed score profiles of Cont., Disc., and Intermt. appear only in 6 cases, or 3.3 percent of the group, 5 of these cases belonging to the indoor experiment.

(8) A method for group testing of auditory localization has been devised, which permits obtaining group averages and corresponding ratings of individual observers, according to scores and score profiles.

(9) It is suggested that the clear *off-on* response of the cortex to intermittent or discrete tones represents an important bio-electrical component of the higher integrative processes, such as auditory space perception.

(Manuscript received December 1, 1941)

References

1. Fernberger, S. W. *et al.*, Perception, *Psychol. Bull.*, 1941, **38**, 432–468.
2. James, H. E. O. and Massey, M. E., Some factors in auditory localization, *Med. Res. Counc., Special Report*, London, 1932, No. 166, 31–51.
3. Marx, W. and Marx, H., Über die Wahrnehmung der Schallrichtung, *Beitr. Phys. Path. u. Therap. d. Ohr., Nase, u. Hals*, 1921, **16**, 32–46.
4. Stevens, S. S. and Davis, H., *Hearing, Its Psychology and Physiology*, New York: Wiley, 1938.
5. Stevens, S. S. and Newman, E. B., The Localization of actual sources of sound, *Amer. J. Psychol.*, 1936, **48**, 297–306.
6. Wallach, H., The role of head movements and vestibular and visual cues in sound localization, *J. Exp. Psychol.*, 1940, **27**, 339–368.
7. Webb, E., Character and intelligence, *Brit. J. Psychol., Monog. Suppl.*, 1915, **1**, No. 3, ix + 99.
8. Wilska, A., Untersuchungen über das Richtungshören, *Act. Soc. Med. 'Duodecim,'* 1938, Ser. A. 21.
9. Woodworth, R. S., *Experimental Psychology*, New York: Henry Holt, 1938.